Mapping RNA Binding Surfaces on Hfq Using Tryptophan Fluorescence Quenching

by

Kirsten Else Hoff

Department of Biochemistry
Duke University

Date:_____

Approved:

_____
Richard G. Brennan, Supervisor

_____
Jack D. Keene

_____
David C. Richardson

_____
Pei Zhou

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Biochemistry in the Graduate School
of Duke University

2013

ABSTRACT

Mapping RNA Binding Surfaces on Hfq Using Fluorescence Quenching

by

Kirsten Else Hoff

Department of Biochemistry
Duke University

Date:_____

Approved:

_____
Richard G. Brennan, Supervisor

_____
Jack D. Keene

_____
David C. Richardson

_____
Pei Zhou

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the Department of
Biochemistry in the Graduate School
of Duke University

2013

Abstract

Hfq is a pleiotropic posttranscriptional regulator and RNA chaperone that facilitates annealing of trans-encoded sRNA/mRNA pairs. It regulates many different cellular pathways including environmental stress responses, quorum sensing, virulence and maintenance of membrane integrity. Hfq is a member of the Sm/LSm family and forms a homohexamer that has two faces, termed proximal and distal. Hfq preferentially binds A/U rich regions that are near stem loop structures. Crystal structures have shown that poly-A sequences tend to bind the distal face while poly-U sequences bind the proximal face. Currently crystal structures reveal the binding mechanisms for short RNA sequences however; physiologically relevant RNA sequences are typically longer and more structured. To study how these more complex RNA sequences interact with Hfq, a tryptophan fluorescence quenching (TFQ) assay has been developed. Here it is presented that TFQ can correctly identify the binding face for two control sequences, A_{15} and U_6, using the *E. coli*, *S. aureus* and *L. monocytogenes* Hfq homologues. Using fluorescence anisotropy and crystallography it is observed that Trp mutants necessary for TFQ may affect binding to some degree but do not affect the overall structure or RNA binding function of Hfq. TFQ is then used to examine the distal face binding motifs for both Gram-negative (*E. coli*) and Gram-positive (*S. aureus*/*L. monocytogenes*) Hfq, $(A\text{-}R\text{-}N)_n$ and $(R\text{-}L)_n$ respectively. Using sequences that

either fulfilled just $(A\text{-}R\text{-}N)_n$ or both $(A\text{-}R\text{-}N)_n$ and $(A\text{-}A\text{-}N)_n$ motifs it is shown that the distal face motif for Gram-negative Hfq is the more specific $(A\text{-}A\text{-}N)_n$ motif. Using sequences that either fulfilled just $(R\text{-}L)_n$ or both $(R\text{-}L)_n$ and $(A\text{-}L)_n$ motifs it is shown that the Gram-positive distal face motif can be redefined to the $(A\text{-}L)_n$ motif. Finally TFQ is used to explore autoregulation of *E. coli hfq*. Two identified binding sites located in the 5′ UTR of *hfq* mRNA, site A and site B, were used for TFQ, along with a longer RNA sequence that contains both sites and their native linker, 5′ UTR. TFQ illustrates that the individual sites and the 5′ UTR are capable of binding both faces. Each site appears to prefer binding to one face over the other, suggesting a model for *hfq* 5′ UTR mRNA binding to Hfq where either one or two *hfq* mRNA bind a single Hfq hexamer. In conclusion, TFQ is a straightforward method for analyzing how RNA sequences interact with Hfq that can be utilized to study how longer, physiologically relevant RNA sequences bind Hfq.

Contents

List of Tables

List of Figures

Acknowledgements

Research is rarely performed by a single person; as such I wish to thank the many people who aided me throughout my graduate career. First I wish to thank Dr. Jillian Orans who worked closely with me during the course of my TFQ studies. I especially thank her for guidance in the purification of Hfq and her never ending patience while teaching me crystallography. I also extend my thanks to Dr. Richard Brennan who allowed me to join his lab when my previous lab was no longer capable of supporting me. He provided me with this excellent project and was a superb mentor. Dr. Richard Brennan and Dr. Maria Schumachers' labs also deserve thanks as they provided friendship and guidance while I was a member. Dr. Nam Tonthat deserves thanks for always being willing to help when I had questions about our x-ray home source, processing crystallography data, or the Panvera beacon 2000.

During my time in Dr. Brennans' lab I mentored two undergraduates who both have my thanks. Ms. Ivy Paw and Mr. John Canty spent many hours purifying protein, performing TFQ experiments and calculating binding constants by FP under my guidance.

While at Duke I have had the privilege to utilize several of Duke's core facilities. At the NMR center I wholeheartedly thank Dr. Ron Venters, Dr. Tony Riberio and Dr. Don Mika who taught me how to run an NMR instrument, collect and process data.

They were always willing to help when I had a question and dealt with the strict time requirements my experiments necessitated. At the Duke Macromolecular X-ray center I wish to thank Dr. Charles Pemble who organized all synchrotron time requests and was willing to answer any questions I had. At the Flow Cytometry Center I thank Dr. Michael Cook who taught me the basics of flow cytometry and aided experiment design. Finally, at the Model Systems Genomics facility I thank Dr. Chandra Tucker for her time discussing yeast experiments and providing single knockout yeast strains that I needed.

I would like to thank my previous mentor Dr. Leonard Spicer for his guidance during the early years of my graduate studies. While in his lab I worked closely with Dr. Patrick Reardon and Dr. Anne Marie Augustus. Their advice was always greatly appreciated. I especially thank Dr. Patrick Reardon and his wife Carrie Marean-Reardon whose friendship over the years has been invaluable to me.

The Duke Biochemistry department is a close knit environment and I wish to thank the many students who were always willing to listen and offer advice when my experiments weren't going as planned. This includes: Dr. Cara Froyd and Dr. Tiffany Sabin, who always answered my random yeast questions, Mr. Gary Kapral and Dr. Vincent Chen, who answered questions I had about the various structural programs I used, Ms. Jessica Wojtasjek, Mrs. Jo Anna Capp and many others.

Finally, I wish to thank my family and friends. They have offered support throughout my time at Duke. Dr. Alice Wong and Ms. Jessica Chu have been the best

friends anyone could ever want, always understanding and always willing to drive down for a visit. My husband and best friend in life, Daniel Hoff, has been the greatest support for surviving graduate school that anyone could ask for.

1. Introduction

1.1 Function of Hfq

Escherichia coli (*E. coli*) Host Factor required for phage Qβ protein RNA Replication (Hfq) was originally identified as a factor required for replication of the RNA plus strand of bacteriophage Qβ *(1-3)*. It has since been shown that disruption of the *hfq* gene causes pleiotropic effects in many bacterial species including *E. coli (4)*, *Listeria monocytogenes* (*L. monocytogenes*) *(5, 6)*, *Sinorhizobium meliloti* (*S. meliloti*) *(7)*, *Borellia burgdorferi* (*B. burgdorferi*) *(8)*, *Neisseria gonorrhoeae* (*N. gonorrhoeae*) *(9)* and others. Some of the observed phenotypes include cell length increases, growth rate decreases, and sensitivity to environmental stressors. Hfq is known to be an essential virulence factor in *E. coli, Salmonella typhimurium (S. typhimurium), Pseudomonas aeruginosa (P. aeruginosa), Vibrio cholerae (V. cholerae)* and other bacteria *(10-14)*. Recently Hfq has been observed to be involved in the multidrug resistance mechanisms of *E. coli* and *Salmonella enterica (S. enterica)* serovar typhimurium *(15, 16)*. This resistance is conferred by Hfq regulation of the AcrB drug efflux pump in *E. coli (15)* and the SmvA drug efflux pump in *S. enterica (16)*.

Hfq has also been shown to play a critical role in many different cellular pathways including cellular response to multiple environmental stressors, maintenance

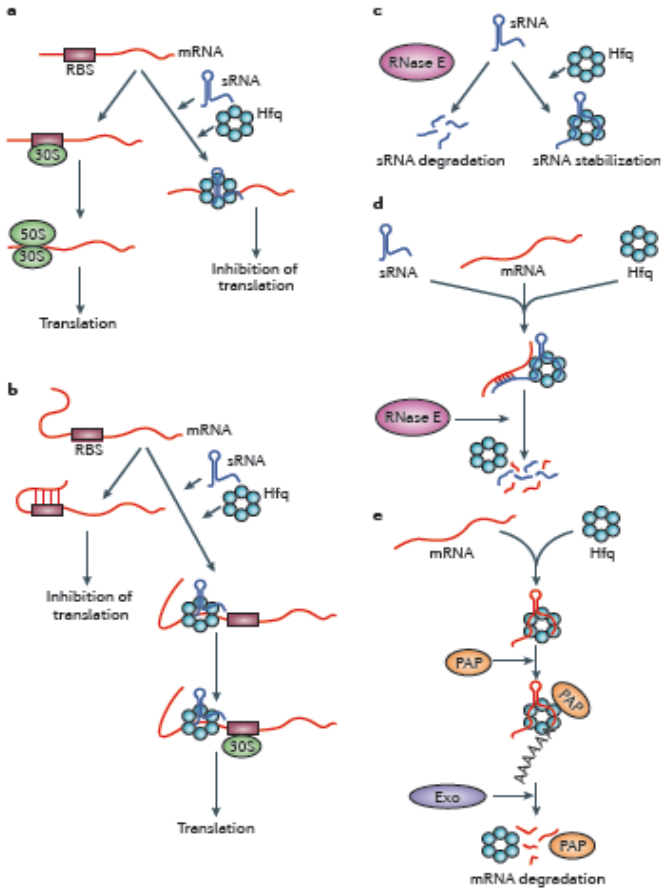


Figure 1: Known Hfq Mechanisms of Action

A) Hfq bound to a sRNA may then sequester the ribosome binding site (RBS) of the target mRNA thus preventing the ribosomal subunits from binding, preventing translation. B) Some mRNA strands have a secondary structure in the 5′ untranslated region (5′ UTR) that blocks the RBS and prevents translation. When Hfq binds to a sRNA it can then bind to the mRNA to prevent formation of the secondary structure thus allowing translation. C) Hfq may protect some sRNAs from being degraded by RNase E by binding to the sRNA. D) Hfq forming a sRNA/mRNA pair may lead to degradation of the sRNA/mRNA pair by RNase E thus leading to downregulation of the mRNA. E) Hfq may bind an mRNA strand and promote polyadenylation by poly(A) polymerase (PAP) which then leads to 3′-5′ degradation by an exoribonuclease thus leading to downregulation. Figure is taken from *(17)*.

of membrane integrity, virulence, and quorum sensing. Hfq functions through controlling expression of targeted genes posttranscriptionally by binding to short stretches of A/U rich regions in small noncoding RNAs (sRNA) and messenger RNA (mRNA). The mechanism of Hfq action usually involves binding a sRNA and facilitating the annealing of that sRNA to its target mRNA. This sRNA/mRNA pairing leads to either upregulation or, more likely, downregulation of the targeted mRNA (Figure 1). There are several different interactions that can happen between sRNA, mRNA and Hfq as can be seen in Figure 1. To upregulate translation, Hfq can bind to a sRNA that pairs with a mRNA and removes a translation inhibitory secondary structure thus allowing ribosomal access to the Shine/Delgarno sequence and promoting translation *(18, 19)* (Figure 1B). Hfq can downregulate translation by binding to a sRNA/mRNA pair in a way that blocks the ribosome from the Shine/Delgarno sequence thus preventing translation (see Figure 1A). It can also bind to a sRNA/mRNA pair which then promotes degradation of the pair by RNase E *(20-22)* (see Figure 1D) or Hfq can bind to an mRNA, promote poly(A) polymerase to polyadenylate the 3′ end of the mRNA sequence which then leads to 3′-5′ degradation of the mRNA by an exonuclease thus leading to decreased expression *(23, 24)* (see Figure 1E). Finally, Hfq can bind to sRNA to sequester it from the cell and prevent degradation of the sRNA which could promote both up and downregulation (Figure 1C)

Facilitating the *trans* sRNA-mRNA pairings has led to Hfq being termed an RNA chaperone. A detailed understanding of how these different mechanisms are influenced by Hfq is essential to understanding how Hfq functions and potentially developing Hfq as a drug target. The number of genes whose expression is affected by Hfq varies between bacterial species. However, it tends to range from 5-25% of all genes within a species *(13, 25-27)*. This large number of genes requires a large collection of sRNA and mRNA sequences be present to ensure proper regulation. Most *trans-encoded* sRNA/mRNA pairs require a helper protein, in bacteria this is most often Hfq. Hfq is at a relatively low concentration compared to the sRNA/mRNA pool concentration *in vivo*. However, the time span between when a sRNA is induced and when major effects are observed from that induction is short at 1-2 min *(20, 28)*.

Hfq has been shown to preferentially bind to A/U-rich strands of RNA. Most sRNA, mRNA and artificial RNA sequences that have been tested for Hfq binding have high affinities, typically ranging from sub- to mid- nM K_D's. As such, a model for how Hfq manages to interact with all these different RNA sequences simultaneously has been hypothesized that is known as the "Active Cycling" model *(29, 30)*. In this model it is proposed that since the Hfq-RNA complex has a low dissociation rate the driver of RNA exchange is the free RNA pool, implying that release of RNA from Hfq follows second order kinetics (Figure 2). Bound RNA contacts several Hfq subunits while a free

competitor RNA contacts a single unoccupied subunit on the same face as the bound RNA. Through a series of reversible steps the competitor RNA can swap with the bound RNA. It has been observed that the dissociation rate constant for an Hfq-RNA

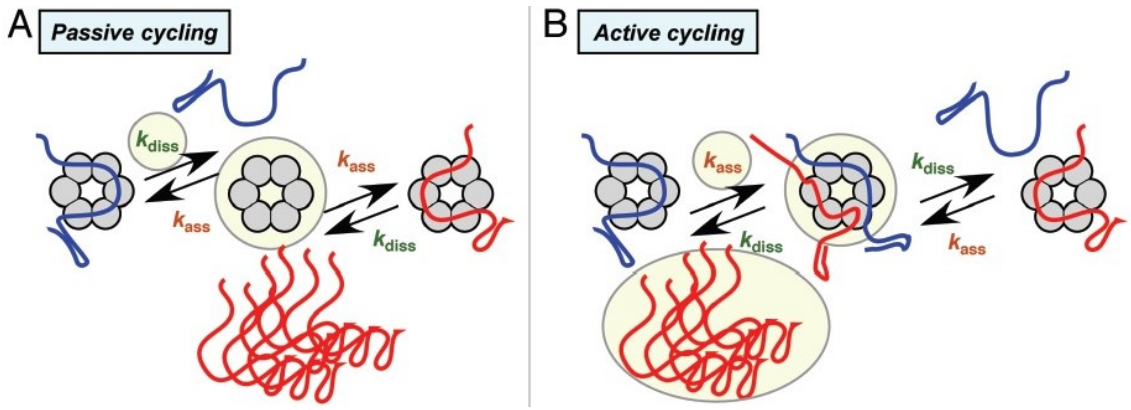


Figure 2: Current Proposed Cycling Models for Hfq

Two current models for how Hfq cycles through different RNA. A) The "passive cycling" model requires that the Hfq-RNA pair dissociate before another free (red) RNA can bind to Hfq. B) The "active cycling" model shows that the free RNA will bind to the Hfq-RNA complex and then potentially lead to dissociation of the bound (blue) RNA, depending on the second order association rate constant and free RNA concentration. Figure is taken from *(30)*.

complex will dramatically increase in the presence of a competitor RNA, dropping from a rate of >150 min down to 1-5 min *(31)*, which adjusts the rate of RNA exchange on Hfq to within the observed *in vivo* rates *(20, 28)*, strongly supporting the active cycling model.

Hfq has been shown to interact with many different sRNA/mRNA pairs including DsrA-*rpoS*, RhyB-*sodB*, Spot42-*galK*, SgrS-*ptsG* and others. These RNA pairs

regulate a variety of different cellular mechanisms. For example the sRNA DsrA binds to the 5′ UTR region of *rpoS* mRNA, thus preventing an inhibitory secondary structure and allowing translation of RpoS to occur (Figure 1B). RpoS regulates stationary phase genes and is important for stress response to environmental factors such as UV radiation, osmotic or temperature shock, oxidative stress, acid or nutrient deprivation. Another sRNA/mRNA pair that Hfq aids formation of is the RhyB/*sodB* pair. RhyB will bind to *sodB* and downregulate the level of *sodB* in the cell. *SodB* encodes superoxide dismutase, which catalyzes the dismutation of superoxide into oxygen and hydrogen peroxide. RhyB is an example of a sRNA that binds and downregulates many different mRNA transcripts via Hfq mediation including the mRNAs for iron storage and iron usage proteins, all mRNAs found in the *sdhCDAB* operon, which encodes succinate dehydrogenase, and others. These pairs are just two examples of the many different sRNA/mRNA pairs that are regulated by Hfq and mediate many cellular functions including stress response, membrane integrity, virulence, drug resistance and quorum sensing. Understanding the structural and biochemical basis of how Hfq manages to chaperone all of these RNA interactions will improve our knowledge of how these pathways function and enable studies on regulating Hfq-RNA interactions for developing as a potential drug target.

1.2 Crystal Structures of Hfq

Hfq is a member of the Sm/LSm family *(32, 33)*. This family of proteins contains both eukaryotic and prokaryotic members. Hfq has been identified in ~50% of all sequenced bacteria and has a highly conserved N-terminus (residues 1-69) while the C-terminal tail is highly variable in both length and sequence (Figure 3). For example, *E. coli* Hfq has a 30 residue C-terminal tail whereas *S. aureus* Hfq only has 9 residues in its C-terminus. The N-terminus has been shown to function without the C-terminus and the role of the C-terminus is still being determined *(34, 35)*.

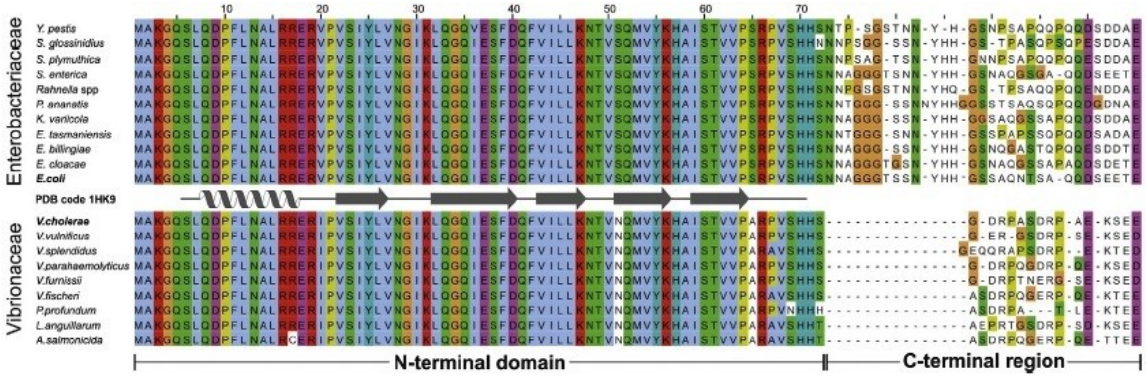


Figure 3: Sequence Alignment of Hfq Homologues

This alignment shows that the N-terminal core is highly conserved while the C-terminus is highly variable in length and sequence. Taken from *(36)*.

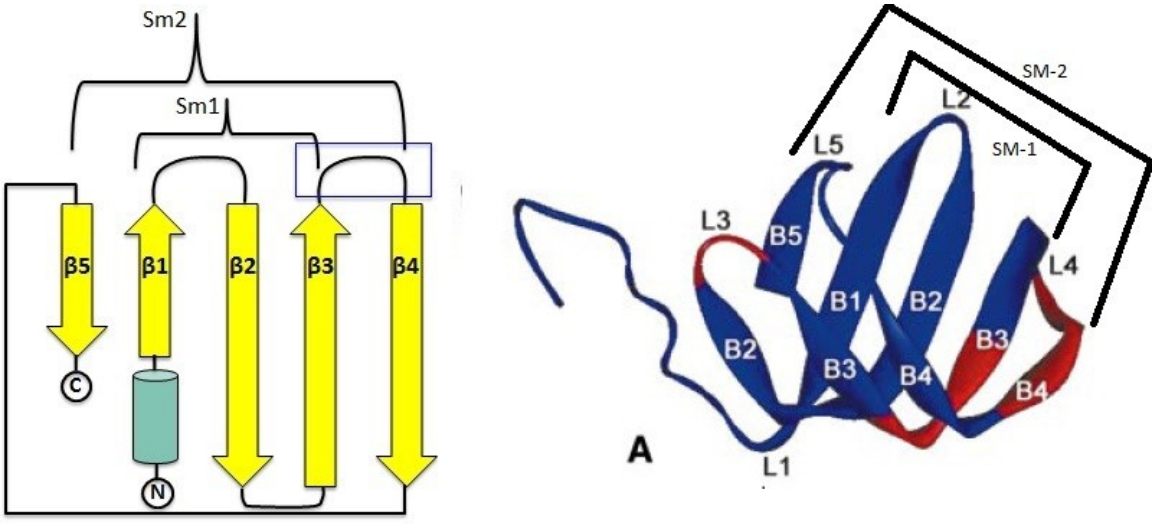


Figure 4: Hfq Greek Key and Monomeric Structure

The secondary and tertiary structure of a Hfq monomer showing the Sm-1, β-strands 1-3, and Sm-2, β-strands 4-5, motifs. Taken from *(37)*.

All members of the Sm/LSm family have a similar monomeric fold that consists of an N-terminal α-helix followed by a five-stranded twisted anti-parallel β-sheet (Figure 4). There are two structural motifs present in the monomer termed Sm-1 and Sm-2. Sm-1 is composed of the first three β-strands while Sm-2 contains the last two β-strands, which are located on opposite sides of the body of the Hfq protomer (Figure 4). The eukaryotic members of the Sm/LSm family typically form a heteroheptameric toroidal shape that has a pore in the middle of the ring. All prokaryotic Sm/LSm family members are Hfq homologues.

The first apo crystal structure of a prokaryotic Hfq protein was that of *S. aureus* Hfq *(38)*. This structure showed that Hfq forms a homohexameric toroidal ring that

orients the N-terminal α-helices all on one face of the ring (Figure 5). Since the α-helices are all found to be on one face, the two faces can be distinguished from each other and

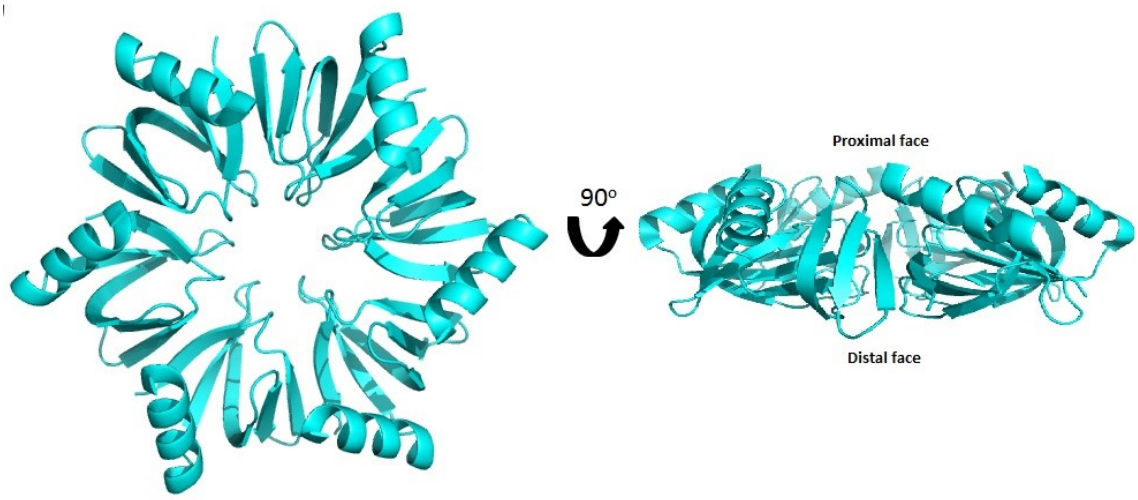


Figure 5: WT *S. aureus* Hfq Apo Crystal Structure

Image made using Pymol from PDB 1KQ1 *(38)*. Hfq is a homohexamer that takes on a toroidal ring structure with all α-helices on one face.

are termed the proximal (α-helix) and distal faces (Figure 5). The *E. coli* Hfq homologue, which also has a described apo structure *(39)*, has a 30 residue C-terminus (compared to *S. aureus* Hfqs' C-terminal tail of 9 residues) however, attempts to observe the C-terminal tail by high-resolution x-ray crystallography have been unsuccessful *(40)*. The observed apo structure for *E. coli* is similar to *S. aureus* with alignment of the hexamers giving an RMSD of 0.90 Å.

There have also been several x-ray crystallography structures solved of Hfq homologues bound to RNA. The first bound structure was of *S. aureus* Hfq-AU_5G *(38)*.

In this structure the RNA sequence was found to be bound in a circular fashion within the pore formed on the proximal face by the α-helices (Figure 6). A single nucleotide is

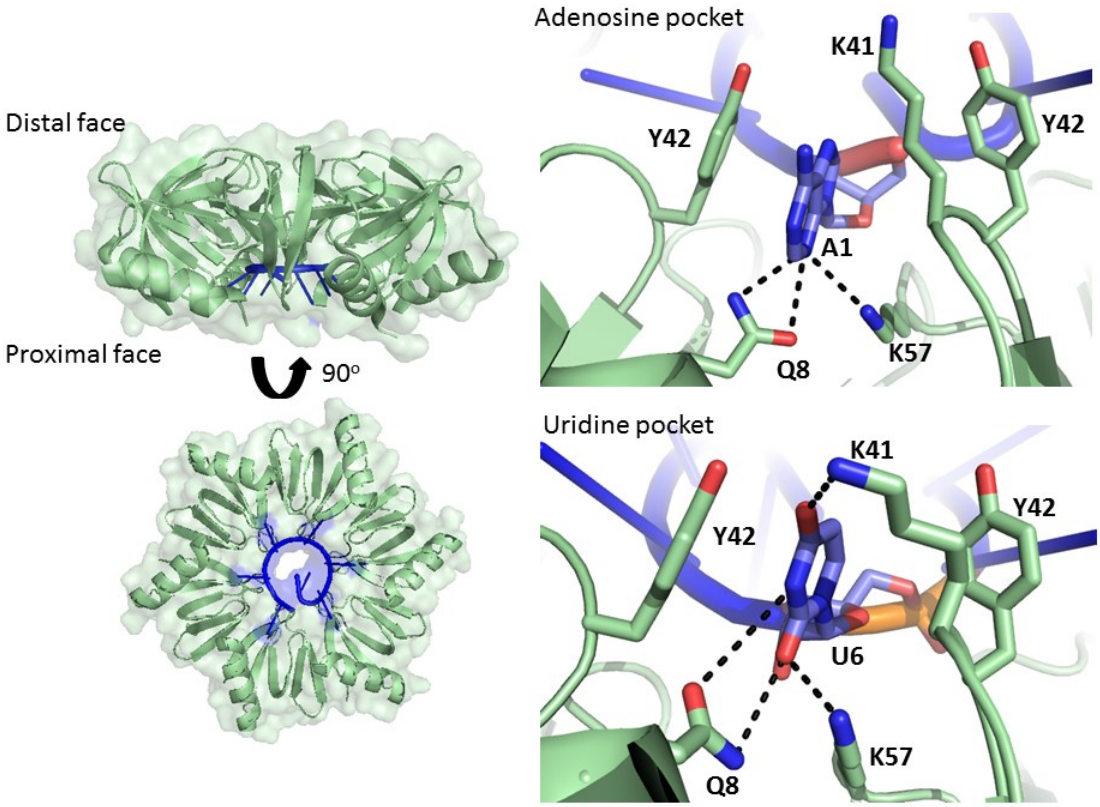


Figure 6: *S. aureus* Hfq-AU$_5$G Crystal Structure

The structure of *S. aureus* Hfq-AU$_5$G. RNA binds within the pore on the proximal face. The pocket that binds to adenosine and a uridine binding pocket are zoomed in to show contacts. RNA is colored blue while Hfq is colored light green. The side chains are colored by atom type with C = light blue (RNA)/green (Hfq), O = red, N = blue. Images made in Pymol using PDB 1KQ2 *(38)*.

shown to be bound per protomer with the adenosine and uridine nucleosides bound while the guanosine points into the pore (Figure 6). The uridines and single adenosine

are shown to be base stacking with residue tyrosine (Y) 42. Cytosine and guanosine appear to be discriminated against by residues lysine (K) 41, K57 and glutamate (E) 8 however these binding sites can accommodate both uridine and adenosine as is seen in the structure (Figure 6). There are three other structures of Hfq homologues bound to RNA on the proximal face, *E. coli* Hfq-AU_6A *(41)*, *E. coli* A_7-Hfq-AU_6A *(42)* and *S.typhimurium* Hfq-U_6 *(43)*. Overall these structures show a similar binding mode to *S. aureus* Hfq–AU_5G with slight differences. They all show base stacking with adjacent aromatic residues and base specificity conferred mainly by hydrogen bonds between the base and either side chain or backbone atoms of nearby residues.

There are also several structures of Hfq bound to RNA on the distal face. These structures include *E. coli* Hfq-A_{15} *(44)*, *E.coli* Hfq-A_7 *(42)*, *E. coli* A_7-Hfq-AU_6A *(42)*, *Bacillus subtilis* (*B.subtilis*) Hfq-$(AG)_3A$ *(45)* and *S. aureus* Hfq-A_4 *(46)*. These structures show two different binding motifs depending on whether the Hfq homologue is from a Gram-negative or Gram-positive bacterium. The *E. coli* Hfq-A_{15} structure shows a circularly bound RNA with three nucleotides bound per protomer subunit in an $(A\text{-}R\text{-}N)_n$ binding motif where A = adenosine, R = any purine nucleoside and N = any nucleoside (Figure 7). The A-site specificity of *E. coli* Hfq is conferred by peptide backbone amide and carbonyl oxygen hydrogen bonding between residue Q33 and the N7 and N6 atoms of adenine and a polar interaction between the Nε amide of residue

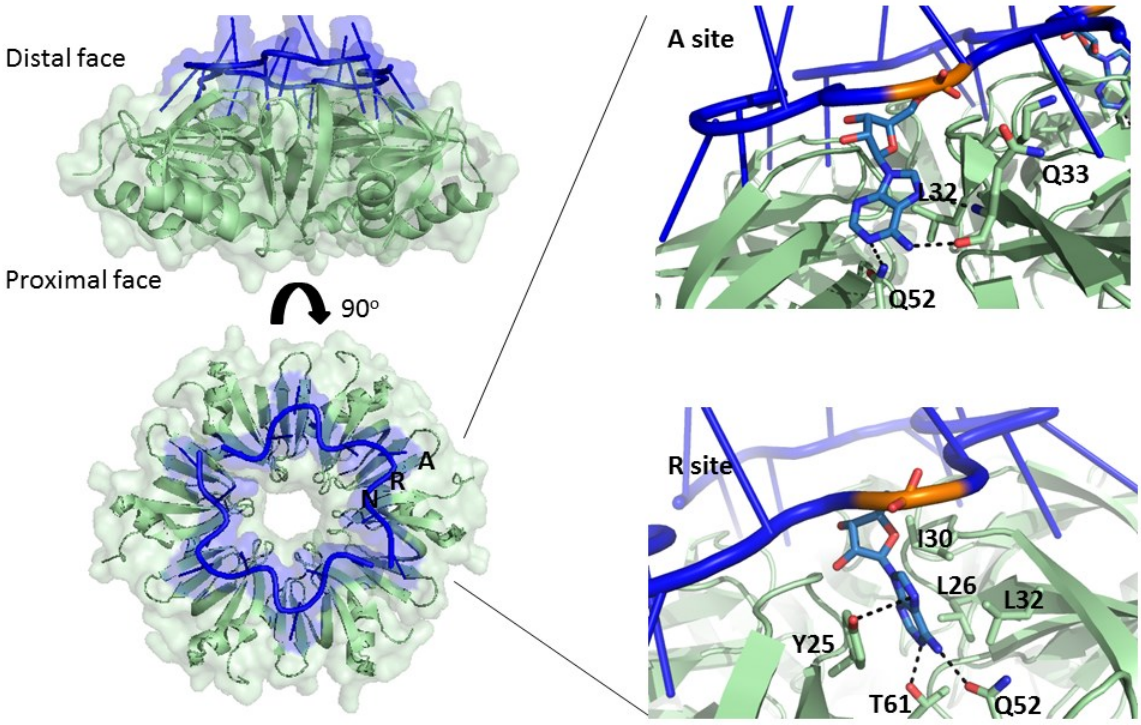


Figure 7: *E. coli* Hfq-A$_{15}$ Crystal Structure

The crystal structure of *E. coli* Hfq-A$_{15}$. RNA binds to the distal face in an (A-R-N)$_n$ motif where A = adenine nucleoside, R = any purine nucleoside, N = any nucleoside. Both the A- and R-sites are zoomed in to show contacts. RNA is colored blue while Hfq is colored light green. The side chains are colored by atom type with C = light blue (RNA)/green (Hfq), O = red, N = blue. The images were made in Pymol using PDB 3GIB *(44)*.

Q52 and the adenine N1 atom (Figure 7). The R-site of *E. coli* Hfq appears to be able to accommodate both adenosine and guanosine whereby the purine ring sticks into a pocket and packs against residues Y25, leucine (L) 26, isoleucine (I) 30 and L32. The adenosine N3, N6, N1 and ribosyl 2' hydroxyl atoms hydrogen bond to Nδ of residue N28, Oε atom of residue Q52, Oγ of residue T61, and the carbonyl oxygen of residue

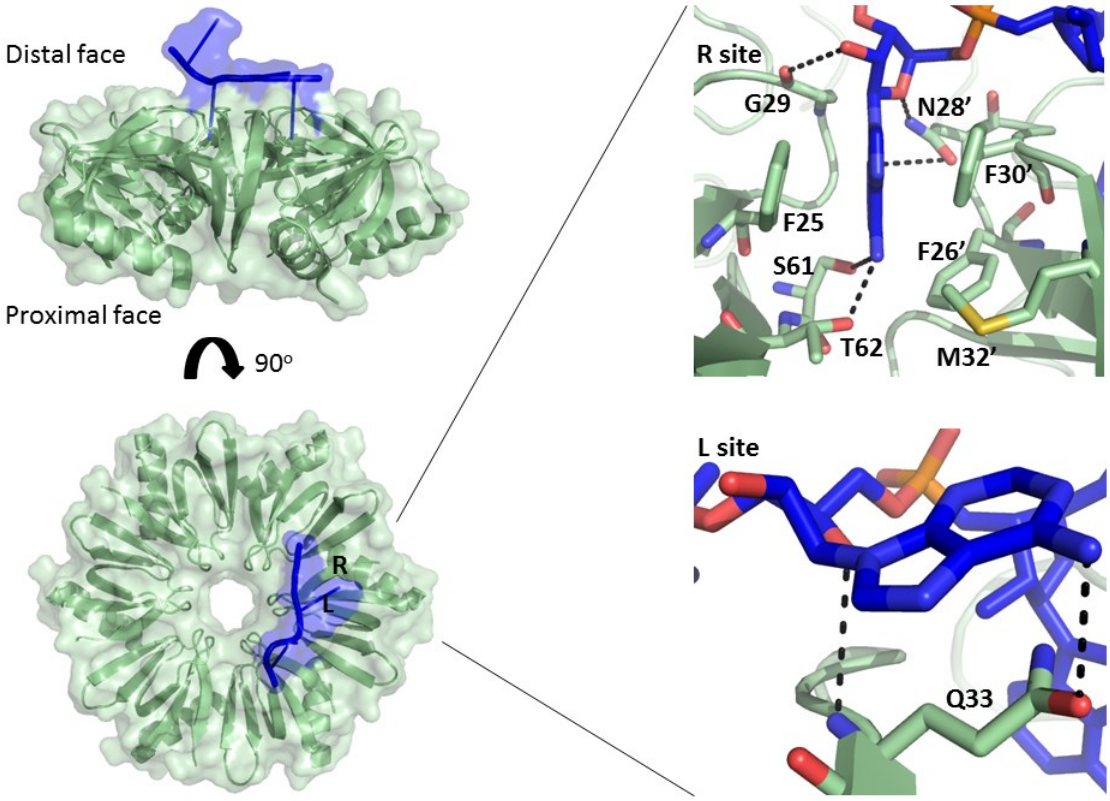


Figure 8: *S. aureus* Hfq-A$_4$ Crystal Structure

The crystal structure of *S. aureus* Hfq-A$_4$. RNA binds to the distal face in an (R-L)$_n$ motif where R = any purine nucleoside, L = R-site linker nucleoside. Both the R- and L-sites are zoomed in to show contacts. RNA is colored blue while Hfq is colored light green. The side chains are colored by atom type with C = light blue (RNA)/green (Hfq), O = red, N = blue. The images were made in Pymol using PDB 3QSU *(46)*.

glycine (G) 29, respectively. The N site has no protein-nucleic acid interactions and likely represents the entrance or exit point for RNA.

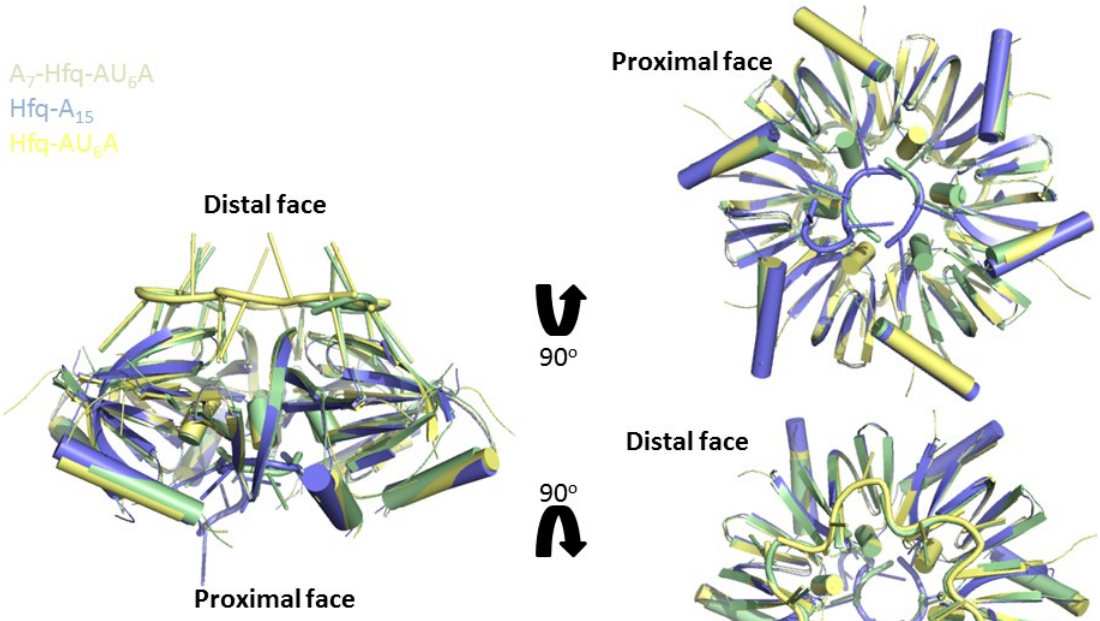
By contrast, the *B. subtilis* Hfq-(AG)$_3$A and *S. aureus* Hfq-A$_4$ complex structures show a significantly different distal-face binding mode, the so named (R-L)$_n$ binding

motif, where R = any purine nucleoside and L = R-site linker nucleoside and can be any base (Figure 8). The R-site is located between β strands 2 and 2′ of adjacent subunits, where the prime indicates the adjacent subunit. The R-site specificity is conferred by base stacking with phenylalanine (F) 25, F26′ and F30′ along with making van der Waals contacts with L27′ and methionine (M) 32′, where the prime indicates the adjacent subunit. There are also selective hydrogen bonds from N28′ to the sugar O4′ and adenine N3 nitrogen, the carbonyl oxygen of G29 to the 2′-OH and between the N1 atom and exocyclic N6 amino group of adenine to the hydroxyl groups of serine (S) 61 and threonine (T) 62. The L-site adenosine base stacks with Q31. The peptide amide of Q31 hydrogen bonds to the sugar O4′. Consequently, the R sites of the *E. coli*, *B. subtilis* and *S. aureus* Hfqs are equivalent but not identical and the mode of protein-nucleobase stacking differs. However, the altered sequences also preclude the formation of the A-site in the *B. subtilis* or *S. aureus* Hfq proteins *(46)*.

On Gram-positive Hfq the $(R\text{-}L)_n$ binding motif presents 12 possible binding sites per hexamer versus the possibility of 18 binding sites for the $(A\text{-}R\text{-}N)_n$ motif in Gram-negative Hfq *(46)*. The presence of an $(A\text{-}R\text{-}N)_n$ or an $(R\text{-}L)_n$ binding mode can be attributed to conserved sequence differences between Hfqs from Gram-negative and Gram-positive bacteria *(46)*.. Consequently, the R sites of the *E. coli*, *B. subtilis* and *S. aureus* Hfqs are equivalent but not identical and the mode of protein-nucleobase

stacking differs modestly. However, the altered sequences also preclude the formation of the A-site in the *B. subtilis* or *S. aureus* Hfq proteins *(46)*.

Recently the first structure of Hfq bound to two RNAs simultaneously has been reported. The solved structure is of WT *E.coli* Hfq bound to A_7 and AU_6A (A_7-Hfq-AU_6A) *(42)*. In this structure it is observed that the overall binding for A_7 is similar to the Hfq-A_{15} structure. Of more interest is the observation that only 4 uridines, rather than the expected six, are found in the expected proximal face binding pockets while the other nucleotides are not resolved in the structure. It is hypothesized that having RNA bound on the distal face causes structural changes that destabilizes binding on the proximal face. However, understanding exactly what those changes are will require further studies as the current structures do not illustrate what has changed to destabilize proximal face binding. An overlay of A_7-Hfq-AU_6A with the corresponding singly bound Hfq-A_{15} or Hfq-AU_6A structures yields RMSDs of 0.62 Å and 0.87 Å respectively (Figure 9). These alignments indicate a high similarity between the structures thus not yielding any obvious explanation for why the bound uridine nucleotide is not seen to bind fully to the hexamer.



	RMSD to A_7-Hfq-AU_6A
Hfq-A_{15}	0.62 Å
Hfq-AU_6A	0.87 Å

Figure 9: Alignment of A_7-Hfq-AU_6A with Hfq-A_{15} and Hfq-AU_6A

Alignment of crystal structures. A_7-Hfq-AU_6A (PDB ID: 4HT9) to Hfq-A_{15} (PDB ID: 3GIB) yields an RMSD of 0.62 Å while A_7-Hfq-AU_6A to Hfq-AU_6A (PDB ID: 3RER) yields an RMSD of 0.87 Å. RMSD calculations were done using secondary structure matching and Cα atom alignment via SSM superpose *(47)*. Images were created using Pymol *(48)*.

All of these structures provide excellent insight into how Hfq homologues bind various short RNA sequences. However, these structures all use either truncated variants of Hfq (*E. coli*) or have a naturally short C-terminal tail and thus do not provide information on the structure or role of the highly variable C-terminal region of Hfq. Structures using the full length *E. coli* Hfq homologue do not have density for the C-

terminus suggesting that the C-terminal tail is highly flexible for apo Hfq *(40)*. Currently the role of the C-terminal tail is poorly understood with some studies indicating that the C-terminus interacts with RNA *(34)* while other reports indicate that the tail is not important for Hfq function *(35)*. These structures also use RNA sequences that are short pieces of what are typically much larger RNA sequences *in vivo* and therefore the structures may not tell a complete story for how Hfq functions within its native environment.

1.3 Small Angle X-ray Scattering of Hfq

Recently a few small angle x-ray scattering (SAXS) studies on full length Hfq homologues in both apo and RNA bound forms have been performed. These studies provide envelopes for full length apo *Vibrio cholerae* (*V. cholerae*) and apo *E. coli* Hfq *(36)*, *E. coli* Hfq-$DsrA_{34}$ sRNA *(49)* and of *V. cholerae* Hfq-Qrr1 sRNA *(50)*. The full length apo Hfq SAXS envelopes indicate that with six-fold symmetry the protein takes on a six pointed star shape with full length *E. coli* Hfq having longer points than full length *V. cholerae* Hfq as would be expected since *E. coli* has a longer C-terminal tail *(36)*. Modeling the crystal structure of the N-terminal core of *E. coli* Hfq into the envelope shows that the N-terminal core sits in the middle of the star, leaving the points of the star as the location of the C-terminal tail (Figure 10A).

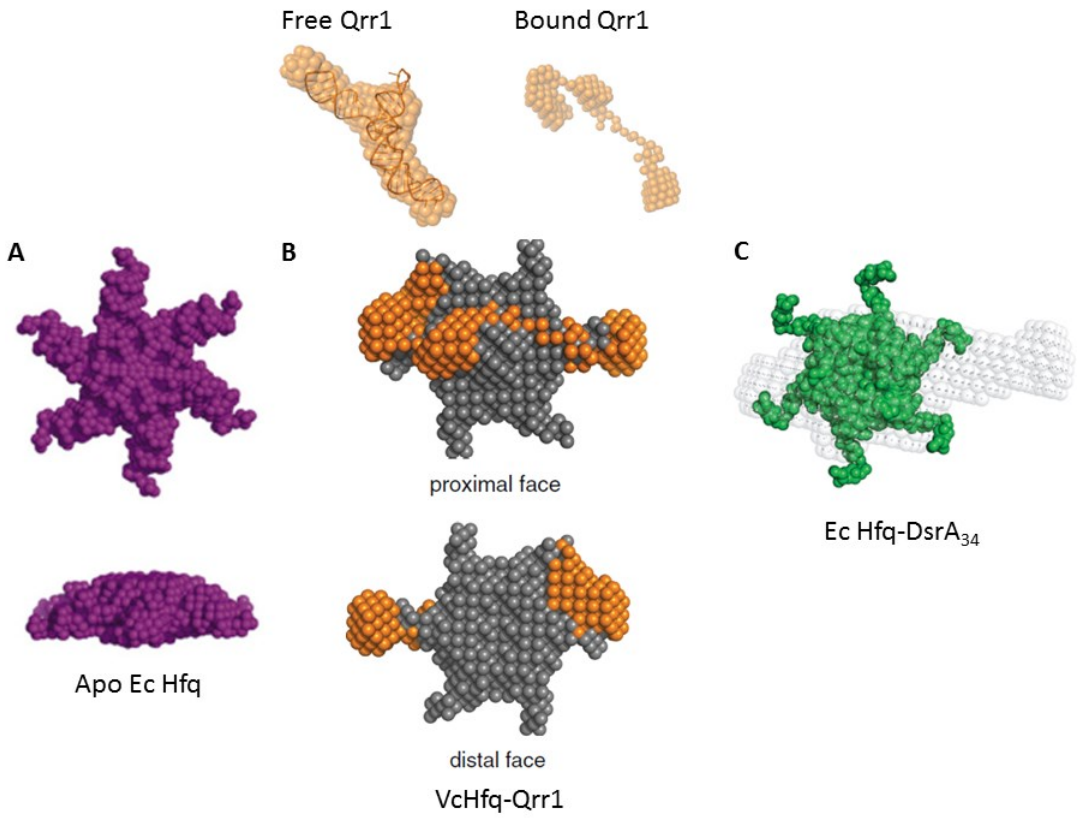


Figure 10: SAXS Envelopes of Apo and Bound Hfq

A) The full length (FL) apo *E. coli* envelope *(36)* shows a six pointed star where the Hfq core resides in the center and the C-terminus is located in the points. The bound envelopes both show RNA bound on one face. B) *V. cholerae* Hfq-Qrr1 complex*(50)* and apo Qrr1. Other experiments were performed to identify the proximal face as the site of Qrr1 binding C) *E. coli* Hfq-DsrA$_{34}$ complex *(49)*. NMR experiments indicate that DsrA$_{34}$ binds on the proximal face. This figure is adapted from figures in *(36, 49, 50)*.

The SAXS envelope calculated for *V. cholerae* Hfq-Qrr1 sRNA shows that Hfq is capable of changing the structure of Qrr1 upon binding and that a single Qrr1 sRNA is bound to an Hfq hexamer (Figure 10B) *(50)*. The SAXS envelope also shows that the Qrr1 sRNA is bound straight across one face of the Hfq hexamer, which was determined

to be the proximal face based on gel shift assays that were performed using proximal and distal face null mutants, which are the point mutants K56A and Y25D, respectively and are designed to knockout RNA binding on that face *(50)*. This result is particularly interesting in that the Qrr1 sRNA does not appear to bind in a circular manner on the proximal face as has been shown to be the case in all x-ray structures of Hfq bound to RNA on the proximal face. This SAXS envelope also shows some potential interaction between the c-terminal residues of Hfq and the Qrr1 RNA suggesting that the presence of the c-terminus may play a role in RNA binding *(50)*.

The final calculated envelope is for *E. coli* Hfq-DsrA$_{34}$ sRNA *(49)* which shows a single RNA strand bound to one Hfq hexamer. This envelope shows DsrA$_{34}$ binding in a non-circular manner on one face of the Hfq hexamer. Through nuclear magnetic resonance (NMR) spectroscopy ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) and ^{1}H-^{13}C HSQC titration experiments it was determined DsrA$_{34}$ binds the proximal face. Specifically, three clusters of residues were observed to experience a change in their environment: 1) the YKH motif (residues 56-58) located in the central pore of the proximal face and the neighboring residues in the adjacent β-strands, 2) the N-terminal residues 2-4 which became structured extending the α-helix on the proximal face, and 3) residues 34-37 that are located in the loop following the β2 strand *(49)*. These NMR observations are in agreement with mutational studies that have also indicated that

DsrA binds to the proximal face *(51)*. This study concludes that the C-terminus of *E. coli* Hfq does not interact with $DsrA_{34}$, which is in contrast to what was observed for *V. cholerae* Hfq-Qrr1.

These SAXS studies show that the full length apo structure of Hfq forms a six pointed star with the N-terminal core located in the middle of the envelope and the C-terminus extending out the points. They also show a 1:1 RNA:Hfq hexamer binding pattern and that the bound RNA does not bind in a circular manner as would be suggested by the known crystal structures. This disparity can easily be explained by the differences between the RNA sequences used to obtain the crystal structures versus the sequences used for SAXS. The RNA sequences in the crystal structures are much shorter sequences ranging from 6-8 nucleotides for binding to the proximal face and 7-15 for binding to the distal face while the SAXS sequences are 34, $DsrA_{34}$ and 99, Qrr1, nucleotides long. The two bound SAXS envelopes reach opposing conclusions on the role of the C-terminus which may be due to differences between the two Hfq homologues or may be due to differences in the RNA sequence lengths. This disparity suggests that further study on the role of the C-terminus is necessary. It is also noted the SAXS studies of Hfq bound to RNA both used sRNA sequences and found binding on the proximal face. It is of interest to also study mRNA interactions with Hfq and

determine if physiologically relevant mRNA sequences bind the distal face as is hypothesized.

1.4 Outstanding Questions to be Answered

Hfq has many different pathways that it regulates through its RNA chaperone activity. It is well known that Hfq functions by facilitating the annealing of sRNA/mRNA pairs. While this function is well known, how Hfq facilitates RNA annealing, while studied for short RNA sequences, is not well understood for longer, more physiologically relevant RNA sequences. The aforementioned crystal structures all illustrate how Hfq interacts with short RNA sequences, which do come from physiologically relevant RNA sequences but do not necessarily show how a full length *in vivo* RNA sequence binds Hfq. The SAXS envelopes also provide information about how an RNA sequence interacts with Hfq. However, SAXS envelopes are low resolution and cannot provide detailed interaction information. In order to know which face an RNA sequence is binding to in a SAXS envelope other experiments must be conducted, such as NMR titrations or gel shift assays. Thus while there is now strong evidence that Hfq can bind RNA on its two faces, along with a third lateral binding site that has been identified only by size exclusion chromatography *(52)*, there is a dearth of information on how Hfq binds to longer, physiologically relevant RNA sequences. Are these longer sequences capable of wrapping around Hfq thus leading to the observed structural

changes? Does the C-terminal tail of Hfq interact with RNA and, if it does, is that interaction important? Currently the role of the highly variable C-terminal tail is poorly understood with some data indicating that the tail does interact with RNA *(34)* while other data indicate that the tail is not important for Hfq function *(35)*. Are the current distal face binding motifs correct?

To address some of these questions tryptophan fluorescence quenching (TFQ) methodology has been developed. TFQ utilizes the fact that *E. coli*, *S. aureus* and *Listeria monocytogenes* (*L. monocytogenes*) Hfq homologues do not have native tryptophan (Trp). Trp fluoresces and using structure-guided design it is possible to place Trp mutants throughout Hfq. RNA will quench the Trp fluorescence of these Hfq Trp mutants, thus addition of RNA that interacts with the Trp mutant will cause a loss of fluorescence, known as TFQ. TFQ correctly identifies the binding surfaces for control sequences A_{15} and U_6. Using TFQ the proposed distal face binding motifs for Gram-negative and Gram-positive bacteria are examined and have been determined to be refined from $(A\text{-}R\text{-}N)_n$ and $(R\text{-}L)_n$ motifs to $(A\text{-}A\text{-}N)_n$ and $(A\text{-}L)_n$ motifs, respectively. This restriction of the distal face binding motif has been previously hypothesized *(46)* due to the fact that all relevant crystal structures have an adenosine bound in the R-site. However, to date no technique has been able to definitively demonstrate that the relevant distal face binding motifs are the more restricted $(A\text{-}A\text{-}N)_n$ and $(A\text{-}L)_n$ motifs. Finally, using TFQ it

is shown that *E. coli* Hfq binds to the previously identified sites from *hfq* mRNA. These two sites are both capable of binding to both faces of Hfq with each RNA binding site demonstrating a preference for one face. A longer sequence (64 nt) from *hfq* mRNA that contains both binding sites also binds to both faces and is the only sequence observed by TFQ to bind to the recently identified lateral binding site. For many tested RNA sequences TFQ is observed for residues that are located throughout the C-terminal tail indicating that it interacts with RNA.

Thus, herein is described a robust method for mapping the binding location of small and reasonably large RNAs. This approach should find wide applicability with Hfq from multiple species. This approach may also be applicable to studying Hfq-protein interactions as long the proteins of interest do not contain Trp.

2. Tryptophan Fluorescence Quenching Correctly Identifies RNA Interaction Surfaces on Hfq

2.1 Tryptophan Fluorescence and Fluorescence Quenching

Fluorescence occurs when an electron in a molecule is excited by a photon that exactly matches the energy required to excite that electron causing the electron to leave the ground state and move into an excited state. Immediately after excitation the

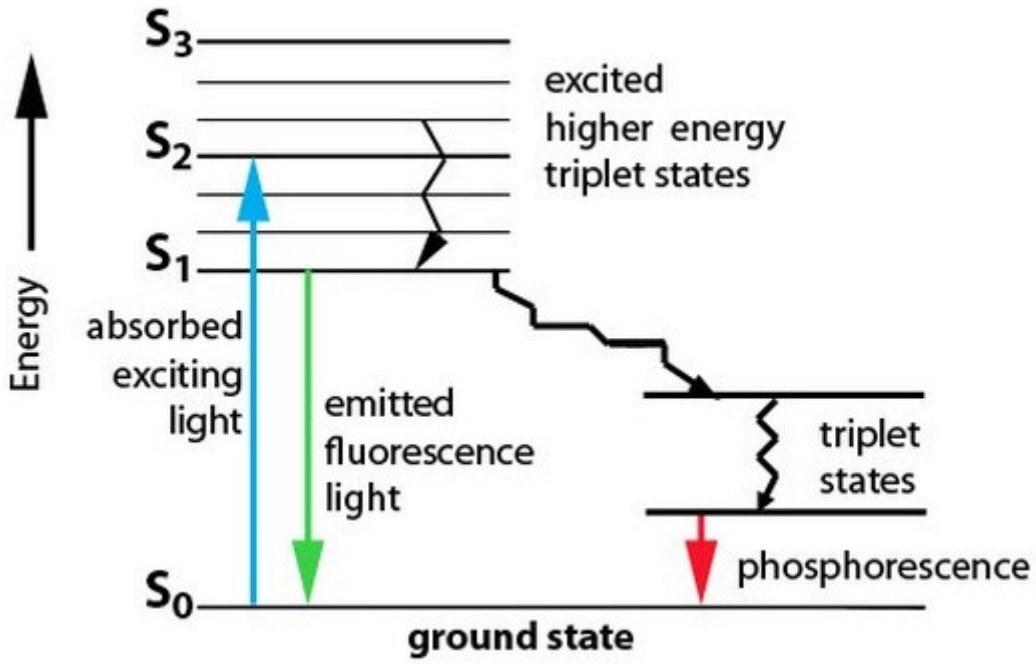


Figure 11: Jablonski Diagram *(53)*.

The Jablonski diagram illustrates how an electron in the ground state is excited by an absorbed photon, moves into an excited state and then releases the absorbed energy as fluorescent light and returns to the ground state *(53)*.

electron releases absorbed energy as either light, heat or to do chemical work and returns to the original ground state (Figure 11). Fluorescence is the emitted light as the electron returns to the ground state and is always at a longer wavelength than the absorbed photon. Many biological molecules fluoresce, including reduced nicotinamide dinucleotide (NADH), chlorophyll, pyridoxal phosphate and proteins *(54)*. Within proteins the aromatic amino acids phenylalanine (Phe, F), tyrosine (Tyr, Y) and Trp (W) are known to be fluorescent. The strongest fluorophore is Trp while Phe has the lowest quantum yield *(54)*. It is possible to selectively excite Trp by using wavelengths above 295 nm as the other aromatic residues are not excited at these wavelengths *(54)*. Fluorescence is sensitive to its environment and will be affected by changes in polarity, salt, pH and for protein fluorescence, ligand binding, subunit association, protein conformational changes, and protein denaturation. Fluorophores have short lifetimes of between 0.5-100 ns however, this short lifetime is sufficient for the excited fluorophore to interact with external molecules that will cause the excited product to release some energy in a nonradiative pathway thus lowering the fluorescence quantum yield *(54)*. This lowering of the quantum yield is known as fluorescence quenching. Tryptophan fluorescence is known to be quenched by iodide, oxygen and nucleotides, to name a few *(55)*.

Hfq does not have any native Trp residues. This lack of Trp allows for Trp residues to be placed in strategic locations throughout the protein. These locations are chosen based on germane structures of Hfq bound to RNA. Using a single Trp point mutant it is then possible to observe the fluorescence of Trp, know where that fluorescence is located within the protein and potentially observe fluorescence quenching by RNA if the RNA sequence interacts with the protein where the Trp residue is located. This technique is known as Trp fluorescence quenching (TFQ).

2.2 Experimental Procedures

Hfq plasmid constructs and site directed mutagenesis – WT *E. coli hfq* full length and truncated (residues 2-69) and WT *S. aureus hfq* full length DNA sequences were provided by Dr. Jill Orans in Dr. Brennan's lab in the pTYB11 expression vector. WT *L. monocytogenes hfq* DNA was ordered from Genscript USA, Inc. (Picastaway, NJ) in the pTYB11 expression vector. All Trp mutants in *E. coli*, *S. aureus*, and *L. monocytogenes* were created using site-directed mutagenesis following the Stratagene site-directed mutagenesis protocol. In brief, this protocol uses primers that contain the mutation desired to PCR amplify DNA, which is then transformed into DH5α cells and plated onto Luria broth (LB) agar selective plates (for the pTYB11 vector these are plates containing 50 μg/mL ampicillin). Colonies are then grown overnight in liquid LB containing 50 μg/mL ampicillin and then miniprepped to purify the DNA vector. The

resulting purified DNA vector is then sequenced (Eton Bioscience, Inc., San Diego, CA) to confirm incorporation of the desired mutation.

Protein Overexpression and Purification – All Hfq homologues and the Trp mutants were transformed into the expression strain ER2566 Δ*hfq*. This strain has the native *E. coli* Hfq deleted so that only the desired Hfq will be purified. To express the protein a 9L LB growth was performed using 50 μg/mL ampicillin for selection. The cells were grown at 37 °C with shaking at 200 rpm until they reached an OD_{600} between 0.4 and 0.6. Protein expression was then induced using 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Once expression was induced the cells were grown for 20 hours at 18 °C with shaking at 200 rpm. Cells were then harvested by centrifugation at 4,000 rpm for 10 minutes (min.) at 4 °C and either stored at -80 °C or lysed immediately. Cell pellets were resuspended in 30 mL lysis buffer (20 mM HEPES pH 8.0, 500 mM NaCl, 0.5 mM EDTA, 1 mM phenylmethanesulfonylfluoride (PMSF), 10 μg/mL DNase, 10 μg/mL RNase). To ensure contaminating RNA and DNA does not come along during purification the crude lysate was stirred at 4 °C for 2 hours before clarification at 17,500 rpm for 30 min at 4 °C. Hfq was then purified from clarified lysate on a chitin column using the IMPACT-CN system as described *(33)*. WT *E. coli* and *S. aureus* Hfq and their Trp mutants were buffer exchanged into 20 mM HEPES pH 8.0, 150 mM NaCl, 0.5 mM EDTA and concentrated to 100 μM-200 μM hexamer. WT *E. coli* and *S. aureus* Hfq was

stored at 4 °C while their Trp mutants were stored at 25 °C due to cold denaturation. WT *L. monocytogenes* and its Trp mutants were concentrated in 20 mM HEPES pH 8.0, 500 mM NaCl, 0.5 mM EDTA to 100 μM hexamer and stored at 25 °C due to cold denaturation.

Crystallization, Data collection, Structure determination and Refinement – All *E. coli* Hfq Trp mutants used for crystallization were C-terminally truncated at residue 69. Hfq Trp mutant crystals were obtained by hanging-drop vapor diffusion method. Drops contained 1 μL protein to 1 μL well solution with a range of protein concentrations (5, 7 and 9 mg/mL). Protein was crystallized from solutions containing 22-28% PEG 3350, 26-32% Isopropanol and 0.1 M Tris pH 8.0-9.0. Crystals were flash frozen in a nitrogen stream at 100 K. X-ray data were collected under cryo-conditions at the Advanced Photon Source (Beamline 22-ID or BM) in Argonne, IL. Data were reduced using HKL-2000 or HKL-3000 *(56)*. Structure determination via molecular replacement was carried out in Phaser *(57)* using PDB ID: 1HK9 as a search model. Structure building and refinement were carried out in Coot *(58)* and Phenix *(59)*, respectively. Selected data reduction and refinement statistics are shown in Table 1. Root Mean Square Deviation (RMSD) calculations were done using secondary structure matching and Cα alignment via SSM superpose *(47)*.

Table 1: Selected Crystallographic Data

	Hfq F11W	**Hfq Y25W**	**Hfq F39W**	**Hfq F42W**
Space Group	$P2_12_12$	$P2_1$	$P2_13$	R3
Cell Dimension				
a,b,c α,β,γ	34.1, 66.3, 84.1 90.0, 90.0, 90.0	31.6, 89.1, 67.0 90, 89.99, 90.0	90.1, 90.1, 90.1 90.0, 90.0, 90.0	104.0, 104.0, 28.2 90.0, 90.0, 120.0
Resolution, Å	50-1.93	50-2.20	50-1.83	50-1.79
I/σ[a]	19.53 (6.01)[a]	51.2 (17.1)	23.42 (2.29)	64.75 (8.49)
R_{sym}, %[a]	6.8 (29.4)	4.1 (8.1)	6.0 (52.4)	5.9 (33.8)
# of reflections	55490	87814	91542	120088
# of unique reflections	15374	18472	21729	10698
Completeness, %[a]	98.2 (90.1)	96.8 (82.5)	99.6 (98.6)	100.0 (100.0)
Refinement				
R_{work}/R_{free} (%)[b]	17.7/21.6	21.3/27.6	20.5/25.7	18.0/23.0
Atoms (#)				
Protein	1623	3059	2145	1048
Solvent	192	61	128	89
Average B Factor ($Å^2$)	28.1	44.5	11.4	34.0
Root-mean-square-deviations				
Bond lengths (Å)	0.007	0.008	0.008	0.005
Bond Angles (°)	1.213	1.110	1.207	1.031
Ramachandran Analysis				
Most favored (%)	95.0	94.6	95.49	96.1
Add Favored (%)	4.0	4.9	2.63	3.9
Disfavored (%)	1.0	0.5	1.88	0.0

Fluorescence Polarization – Fluorescence polarization measurements were performed with a PanVera Beacon 2000 instrument (Invitrogen, Madison, WI, USA) at 295 K. Hfq was serially diluted into 100 μL of binding buffer containing 20 mM HEPES pH 8.0, 0.5 mM EDTA, 200 mM NaCl and 1 nM fluorescein-labeled RNA. Samples were excited at 490 nm and emission was detected at 530 nm. Data were analyzed using Kaleidograph assuming a 1:1 stoichiometry between one Hfq hexamer and one molecule

of RNA. The data were plotted using KaleidoGraph (Synergy Software) and the generated curves were fit using non-linear least squared analysis, assuming a bimolecular model such that the K_d values represent the protein concentration at half maximal RNA binding *(60)*. The binding isotherms were fit to the equation, $P = [(P_{bound}-P_{free})\ [protein]/(K_d+[protein])]+P_{free,}$ with P_{bound} being the maximum polarization at saturation, P is the polarization at a given protein concentration, P_{free} is the polarization of free fluorescein-labeled RNA and K_d is the equilibrium dissociation constant. All values were independently determined in triplicate.

Tryptophan Fluorescence Quenching (TFQ) – Fluorescence quenching measurements were performed using an RF-5301PC spectrofluorophotometer (Shimadzu, Nakagyo-ku, Kyoto, Japan) at 298 K. Fluorescence quenching was done by exciting Trp at 298 nm and scanning the emission fluorescence from 320-400 nm. A 1 mL sample containing 1 μM of each Hfq protein in binding buffer (20 mM HEPES pH 8.0, 150 mM or 200 mM NaCl, 0.5 mM EDTA) was scanned and then a titration was done using a specific RNA sequence. Two concentration ranges were examined; initially 10 – 40 μM RNA to determine the overall sensitivity of the system and then 1 - 4 μM RNA in an attempt to decrease ambiguity that might result from secondary, lower affinity binding. Each titration was done in triplicate. Data were analyzed using Microsoft Excel. Quenching was determined using the arbitrary fluorescence maximal height for

each Trp mutant. Quenching percentage was calculated using the following equation: (1 – ((F_R-F_B) ÷ (F_0-F_B))) x 100, where F_R is the fluorescence value after addition of RNA to the Hfq solution, F_0 is the initial fluorescence value of the Hfq solution without RNA and F_B is the fluorescence of buffer without RNA or Hfq. All data were adjusted to exclude buffer effects.

RNA samples and generation of the 5'-UTR of hfq mRNA – Oligoribonucleotides were purchased from IDT (Coralville, IA) or Oligos etc (Wilsonville, OR) and used without further purification. The 64 nucleotide 5'-UTR of the *hfq* mRNA of *E. coli* was synthesized biochemically. Briefly, the pMCSG7 vector was modified to include the DNA sequence 5' - TAA TAC GAC TCA CTA TAG GAT TTT TTC AGA ATC GAA AGG TTC AAA GTA CAA ATA AGC ATA TAA GGA AAA GAG AGA ATG GGA TCC – 3' using standard ligation independent cloning (LIC) techniques. To prepare RNA, the modified vector was midiprepped, phenol:chloroform extracted, and 8 to 20 μg of clean DNA vector was digested using BamH1-HF (NEB, Ipswich, MA) at 37 °C for 16 hours. The T7 RNA polymerase reaction (100 μL 10X Buffer (0.5 M Tris pH 7.5, 0.25 M $MgCl_2$, 0.05 M EDTA), 40 μL 50 mM spermidine, 200 μL rNTP mix (20 mM/rNTP), 30 μL 100 units/mL inorganic pyrophosphatase (NEB), 80 μL 1 mg/mL T7 RNA polymerase, 60 μL 8-20 μg/mL digested DNA vector, 440 μL dH_2O) was carried out at 37 °C for 16 hours and quenched with 2 mL 0.45 M EDTA. The reaction was then purified by acidic

phenol:chloroform extraction and ethanol precipitated at -80 °C for 16 hours. RNA was centrifuged at 13,000 rpm for 30 min. at 4 °C. The RNA pellet was further purified using 70% ethanol, centrifuged at 13,000 rpm for 30 min. at 4 °C and air dried before resuspending in 100 µL 10 mM sodium cacodylate pH 6.5.

2.3 Tryptophan Fluorescence Can Correctly Identify Known Binding Faces

2.3.1 *Escherichia Coli* Hfq

2.3.1.1 Development of TFQ Assay

Initially four Trp mutants were created in *E. coli* Hfq to verify that a TFQ method is robust and can correctly identify known binding interactions. *E. coli* Hfq residue Y25, which stacks with the R-site bound adenine ring of the Hfq-A_{15} and A_7-Hfq-AU_6A structures on the distal face, was mutated to Trp (Y25W) *(42, 44)*. Similarly, the *E. coli* Hfq-AU_6A RNA structure showed base stacking between residue F42 and uridine, hence the F42W mutant was used to examine proximal face binding *(41)*. We also mutated residue F39 (F39W) as a proximal face "edge" binding site, based on its recently reported importance in binding the sRNAs ArcZ and McaS *(61)*. To explore potential binding to the positively charged cleft on the surface of *E. coli* Hfq we chose to mutate F11 (F11W). All mutations were generated using both the full length WT *E. coli* Hfq and the *E. coli* Hfq C-terminal truncation protein, which contains residues 2-69.

Initially the four truncated mutants were purified and prepared for TFQ. Each mutant was used at 1 μM in a buffer containing 150 mM NaCl. An initial fluorescence scan was taken by exciting Trp at 298 nm and scanning Trp emission from 320-400 nm. Then the desired RNA substrate was titrated into the sample in 10 μM increments until a

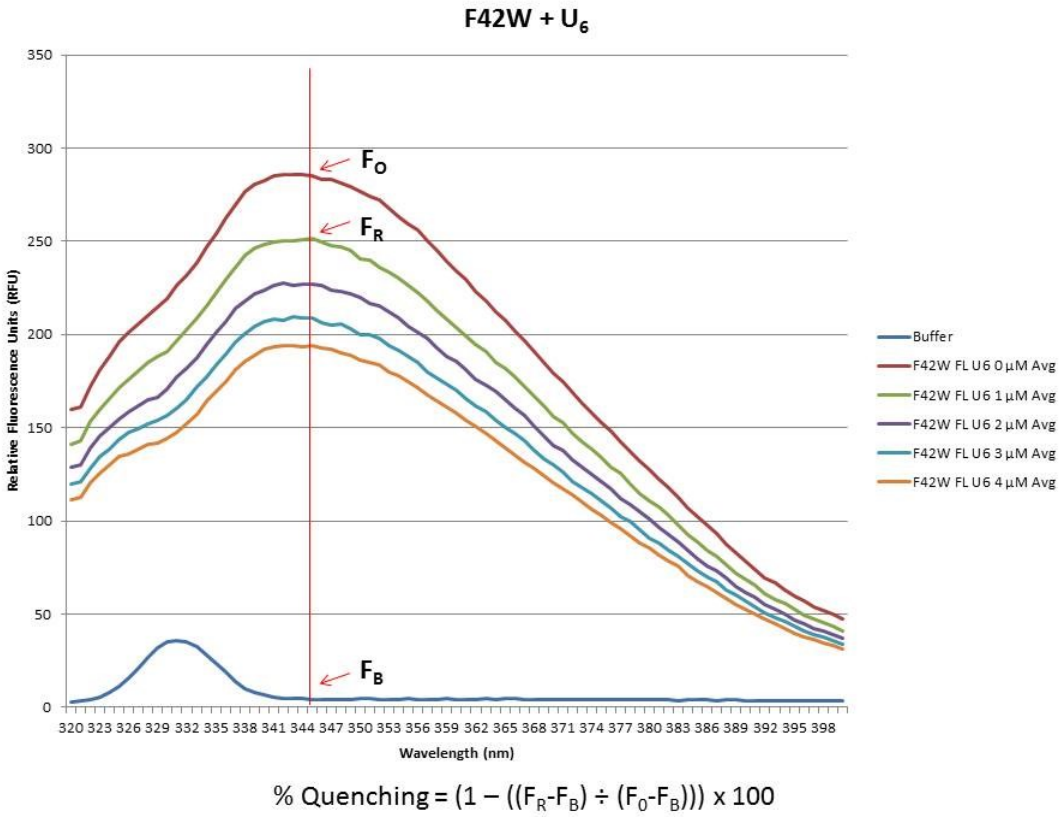


Figure 12: Tryptophan Fluorescence Quenching Example

This graph illustrates the raw data from a Trp quenching experiment when quenching occurs. To determine the percentage of quenching the equation (1 – ((FR-FB) ÷ (F0-FB))) x 100 is used. FR is the fluorescence value after addition of RNA to the Hfq solution, F0 is the initial fluorescence value of the Hfq solution without RNA and FB is the fluorescence of buffer without RNA or Hfq

final concentration of 40 μM was reached (Figure 12). To determine the percentage of quenching the simple equation $(1 - ((F_R-F_B) \div (F_0-F_B))) \times 100$ is used. F_R is the fluorescence value after addition of RNA to the Hfq solution, F_0 is the initial fluorescence value of the Hfq solution without RNA and F_B is the fluorescence of buffer without RNA or Hfq (Figure 12). Each mutant is tested for quenching in triplicate and the quenching results for all mutants are plotted as a bar chart. We define a significant Hfq-RNA interaction when the final titration has quenching that exceeds 10% and is 2-fold greater than the initial titration quench. If both faces meet these criteria then we conclude the RNA substrate is able to bind both faces. If one face is quenched at least 1.5-fold greater than the other, then that face is defined as being the preferential binding face.

The control RNA sequences used were A_{15} and U_6, which bind to the distal and proximal faces, respectively. The results from these initial experiments indicated that some non-specific binding interactions were being detected as the RNA sequence A_{15} was observed to quench both faces, with a preference for binding to the distal face, along with showing some stability issues for both the F11W and F39W mutants (Figure 13). Specifically the F11W and F39W mutants were observed to precipitate out of solution. This precipitation then would cause an artificial increase in fluorescence due to non-specific light scattering.

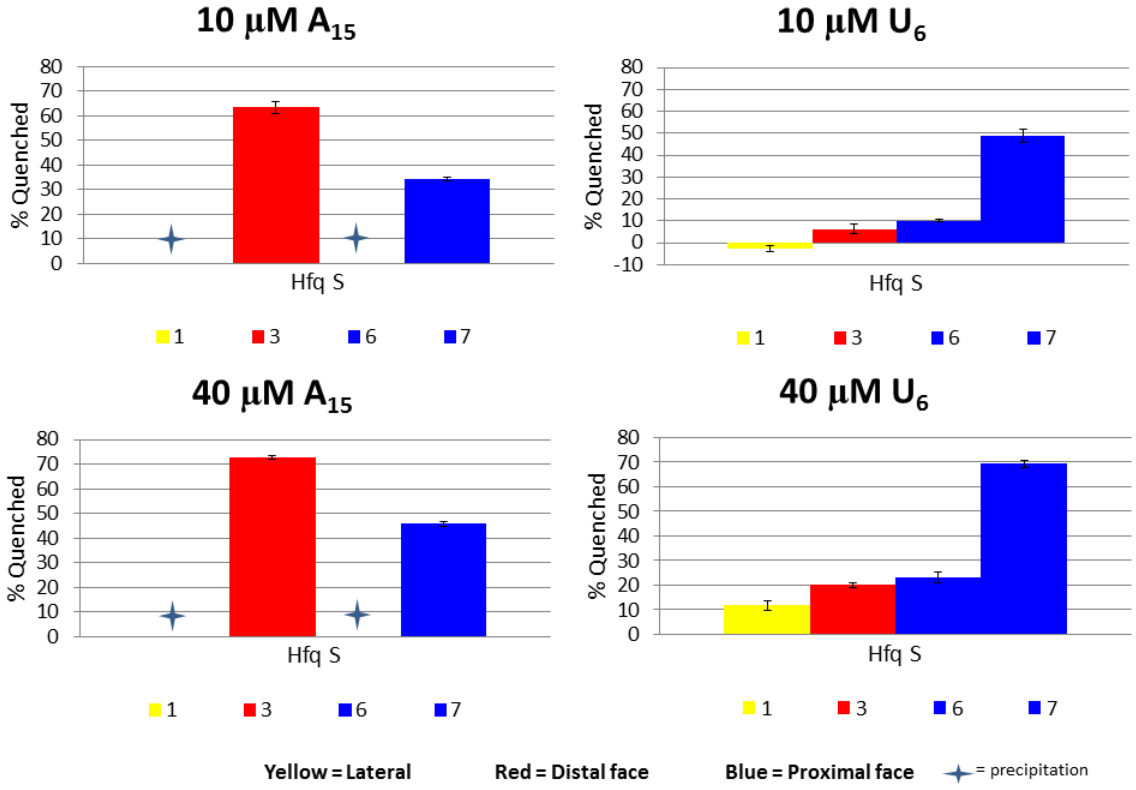


Figure 13: Initial TFQ Experiments Using Hfq Truncated Mutants

Initial control experiments using four point mutants, 1 = F11W, 3 = Y25W, 6 = F39W, and 7 = F42W made in the truncated (residues 2-69) *E. coli* Hfq. These graphs show either non-specific interactions or lower affinity interactions at these RNA concentrations.

To determine if the missing C-terminal residues were involved in the RNA interactions the full length F11W, Y25W, F39W and F42W mutants were created and next used for TFQ experiments titrating RNA from 10 - 40 µM (Figure 14). This experiment indicated that the full length *E. coli* Hfq is more stable in the presence of

RNA as F11W and F39W no longer precipitate, however it did not improve the observation of lower affinity binding interactions for A_{15} to the proximal face.

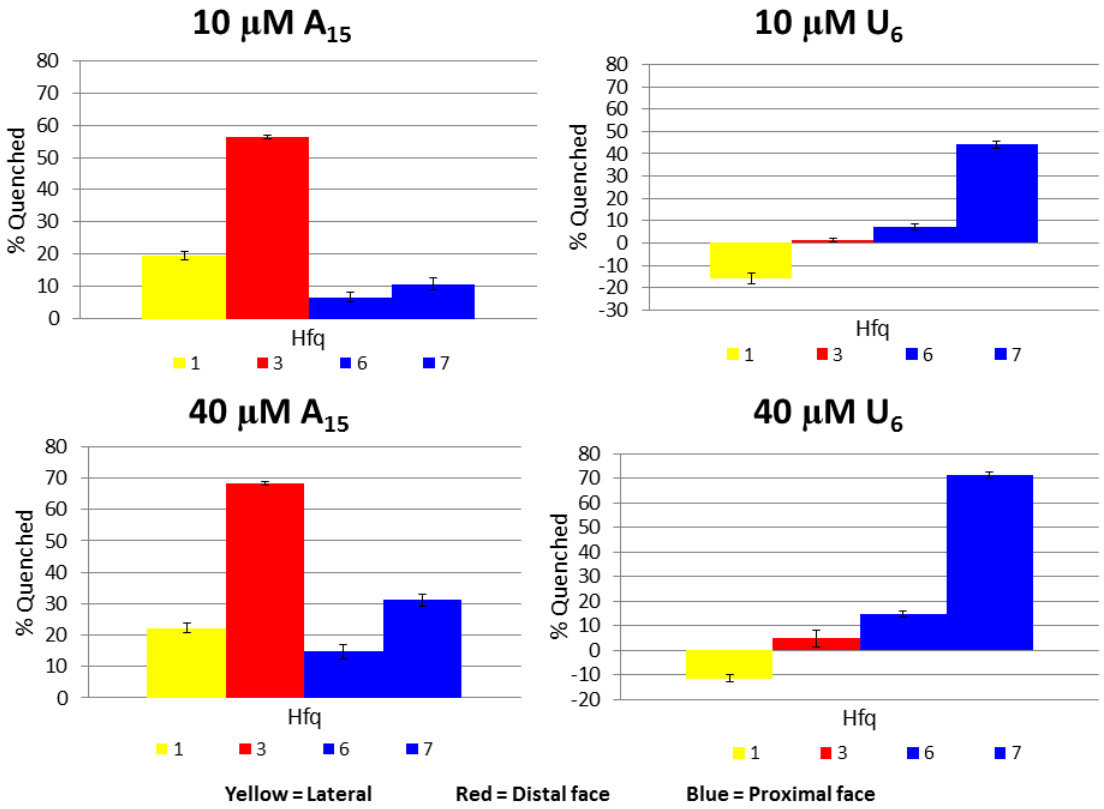


Figure 14: Initial TFQ Experiments Using Hfq Full Length Mutants

Control TFQ experiments using the point mutants 1 = F11W, 3 = Y25W, 6 = F39W and 7 = F42W that were created in full length *E. coli* Hfq. Including the C-terminal tail aided in preventing precipitation however; it did not reduce the observed lower affinity binding of A_{15} for the proximal face.

To control non-specific interactions the salt concentration can be raised. Therefore to reduce nonspecific interactions the NaCl concentration in the binding buffer was increased from 150 mM to 200 mM and the TFQ experiment was repeated

using the truncated *E. coli* Hfq mutants (Figure 15). Raising the salt concentration improved stability of the proteins and significantly reduced the observed lower affinity interactions thus indicating that higher salt concentration should be used. However, the lower affinity binding of A_{15} to the proximal face is still detected.

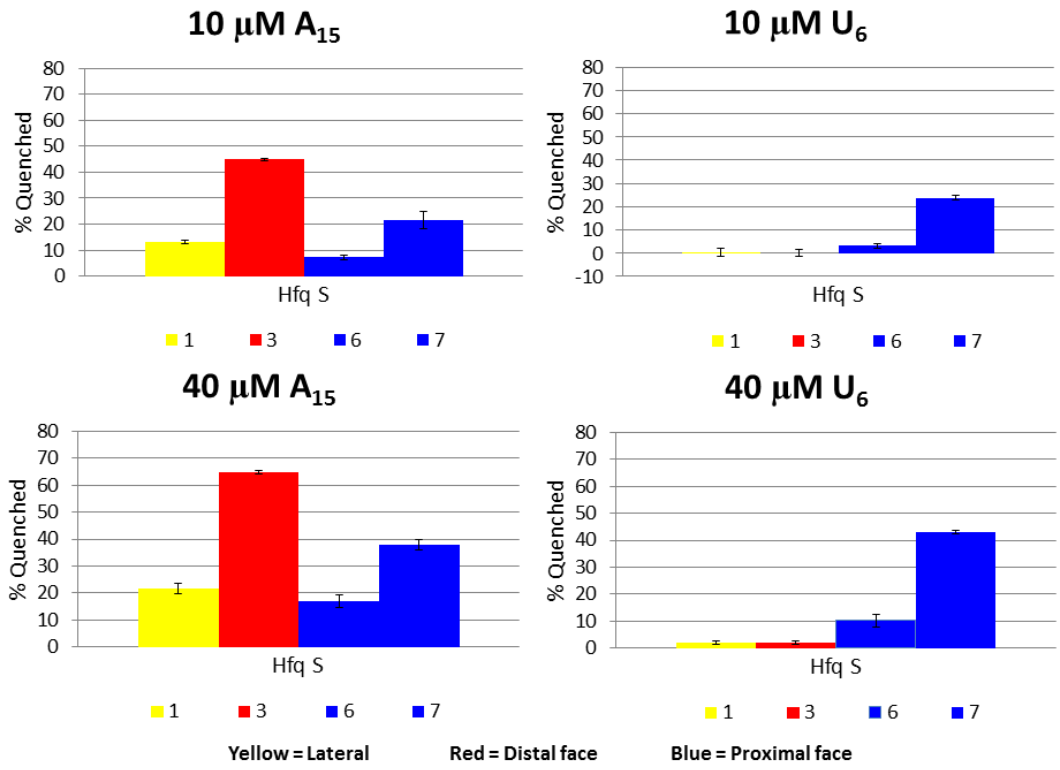


Figure 15: TFQ with 200 mM NaCl and the *E. coli* Hfq Truncated Mutants

Control TFQ experiments using the truncated point mutants 1 = F11W, 3 = Y25W, 6 = F39W and 7 = F42W of *E. coli* Hfq. The sodium chloride concentration present in the binding buffer was raised from 150 mM to 200 mM to reduce nonspecific interactions. Raising salt concentration reduced nonspecific interactions indicating that higher salt concentrations should be used for all TFQ experiments.

Finally, as another step in eliminating lower affinity binding and thus quenching of less preferred sites, the RNA concentration was lowered by 10-fold and a range of 1-4 μM was used to see if the lower affinity interactions could be further reduced using a

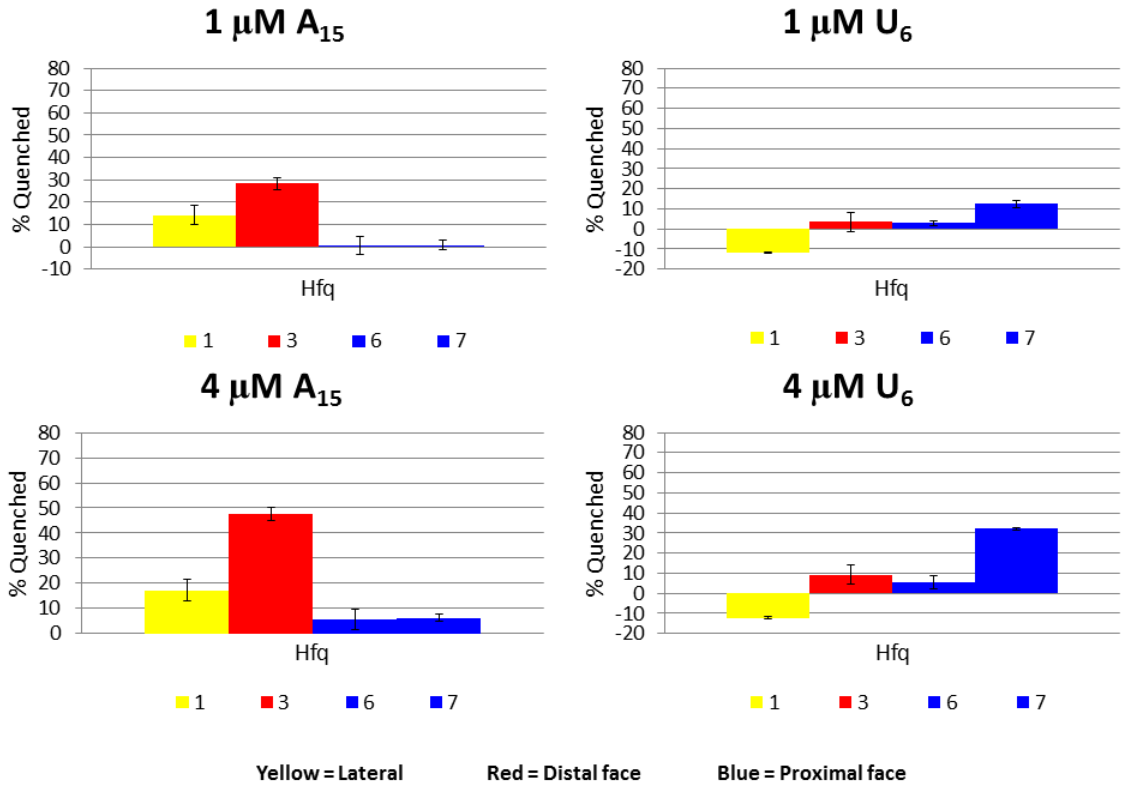


Figure 16: TFQ using 10-fold less RNA with Full Length Hfq Mutants

Control TFQ experiments using the Trp point mutants 1 = F11W, 3 = Y25W, 6 = F39W and 7 = F42W that were made in the truncated (residues 2-69) *E. coli* Hfq. The RNA concentration range was reduced 10-fold to bring the ratio of protein:RNA into a more appropriate range (1:1 – 1:4). Reducing the amount of RNA titrated into the protein sample dramatically reduced the observed lower affinity interactions while not significantly impacting expected binding.

concentration range that put the Hfq hexamer:RNA ratio into the 1-4 range versus the 10-40 range (Figure 16). This drop in the RNA concentration did reduce the lower affinity interactions and at the same time did not significantly impact the quenching of those residues that are expected to interact. Thus the design for the TFQ experiment was finalized to use the full length Hfq sequence, 200 mM NaCl in the binding buffer and a range of 1 to 4 μM RNA.

2.3.1.2 TFQ controls for *E. coli* Hfq

It was observed that the distal face mutation, Y25W worsened A_{15} binding by 180-fold (see section 2.4). While this significant drop in binding did not prevent binding of A_{15} it is clear that weaker binding RNA substrates may not bind to the Y25W mutant and no quenching would be observed hence resulting in a mixed binding site or incorrect identification of the binding site. Therefore two other distal face Trp mutants were designed based on the *E. coli* Hfq-A_{15} and A_7-Hfq-AU_6A structures, K31W and Q33W. K31 is about 6 Å from bound RNA on the distal face of *E. coli* Hfq and does not appear to interact, thus creating a negative control for recognizing the distal face binding motif while Q33 interacts with A_{15} through its backbone carbonyl and amide groups. Modeling Trp in the Q33 position suggests that the Trp side chain may be able to base stack with the bound adenosine or at least approach the adenine ring closely (Figure 17). Another mutation, R17W, was made at this point to study a recently proposed lateral

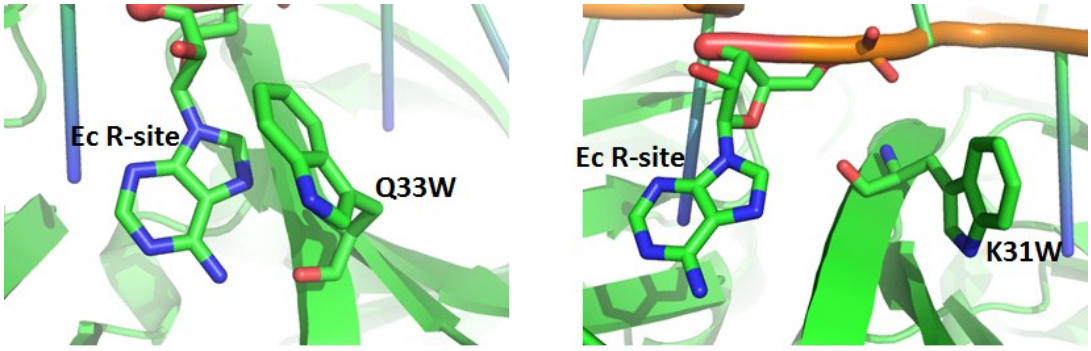


Figure 17: Modeling of Trp into Residues Q33 and K31

Residues Q33 and K31 were modeled as Trp point mutants to analyze the likelihood that the identified target amino acid would base stack with bound ribonucleotide. It is shown that Q33W is capable of base stacking with the bound adenosine while K31W does not thus suggesting that Q33W will experience quenching in the presence of adenosine while K31W will not. Modeling was performed using Pymol *(48)*.

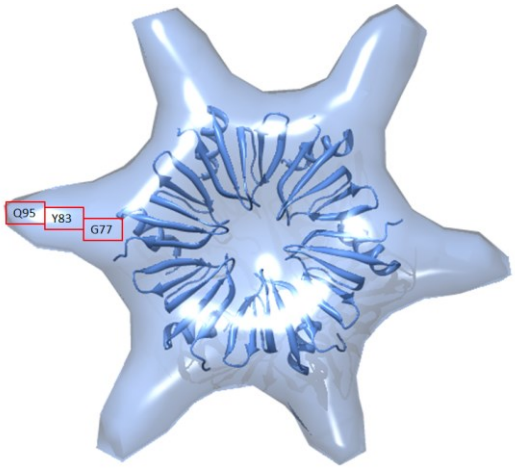


Figure 18: SAXS Envelope of Full Length *E. coli* Hfq Illustrating Location of C-terminal Mutants

A SAXS envelope of full length WT *E. coli* Hfq. The crystal structure of a truncated (residues 2-69) WT *E. coli* Hfq is fit into the envelope to illustrate where the C-terminal tail residues are expected to be. The red boxes indicate the speculative location of the mutated residues.

binding site identified for RhyB RNA binding to St Hfq *(52)*. Finally, to further explore the role that the C-terminal residues may play in RNA binding for *E. coli* Hfq, three mutations were designed throughout the length of the tail; G77W, Y83W and Q95W (Figure 18).

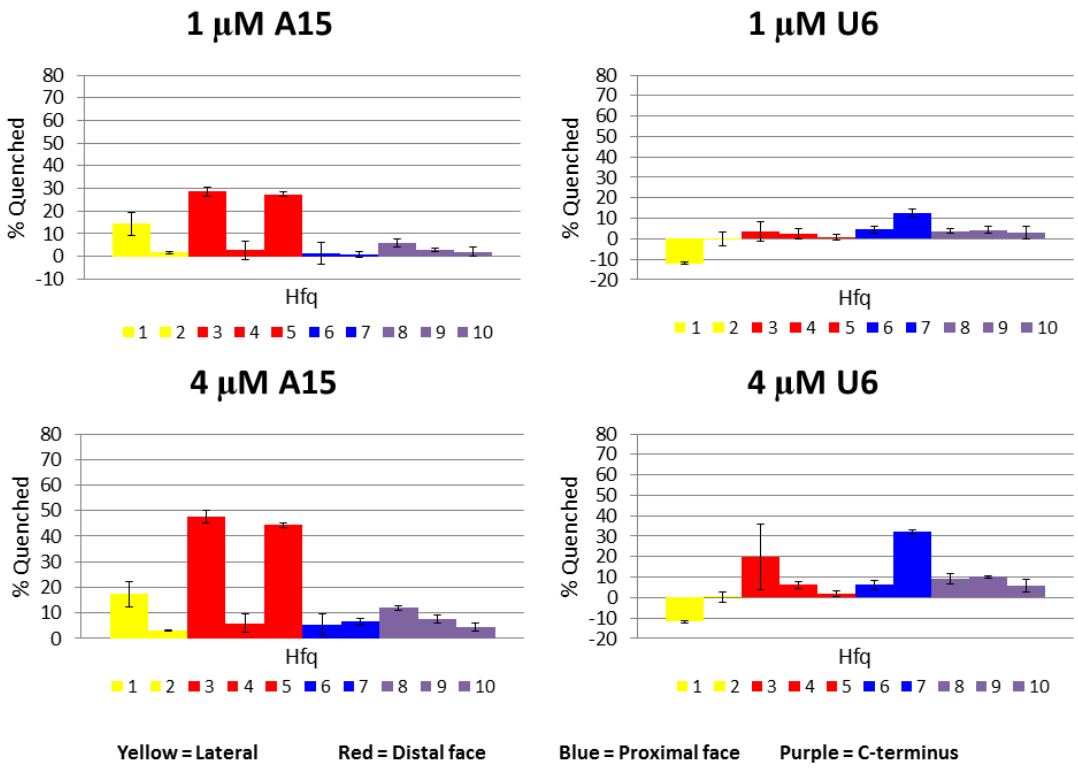


Figure 19: Final *E. coli* Hfq TFQ A_{15} and U_6 Graphs

Control TFQ experiments using the point mutants 1 = F11W, 2 = R17W, 3 = Y25W, 4 = K31W, 5 = Q33W, 6 = F39W, 7 = F42W, 8 = G77W, 9 = Y83W, 10 = Q95W that were made in the full length *E. coli* Hfq. The sequences A_{15} and U_6, which bind the distal and proximal faces respectively, were titrated from 1 to 4 µM. As these graphs illustrate, both sequences were observed to bind to the expected face.

With all of these mutants in hand the designed TFQ experiment was performed in triplicate using the control sequences A_{15} and U_6 (Figure 19). It was observed that TFQ correctly identifies the binding face for a particular RNA. Specifically on the distal face the Y25W and Q33W mutants are significantly quenched at all A_{15} RNA concentrations while the K31W mutant is not quenched as would be expected if the binding mode has not been disturbed. There is some quenching observed for both F11W and G77W indicating that these residues may interact with longer A-tract RNA. Also for A_{15} the proximal face mutants, F42W and F39W, are not quenched in the 1-4 μM RNA range although they are quenched in the 10-40 μM range, indicating some lower affinity binding of A_{15} at high RNA concentrations to the proximal face. Residues R17W, K31W, Y83W and Q95W do not quench in the lower micromolar range indicating no interaction occurring between these residues and A_{15}.

U_6 was also observed to bind as expected, i.e. to the pore of the proximal face, with significant quenching being observed for F42W at all RNA concentrations (Figure 19). F11W has an enhanced fluorescence signal upon U_6 binding suggesting that the environment F11W experiences is being altered, most likely by some slight structural rearrangement. Since the increase in fluorescence happens immediately upon RNA addition and then does not continue to increase this change is unlikely due to RNA binding. No other tested mutants quench in the lower concentration range (1 – 4 μM). In

the 10-40 μM RNA range, residue F39W shows some quenching, although this is not as significant as that observed for residue F42W, but suggests a second, weaker affinity binding site. These results demonstrate that Trp residues show fluorescence quenching when the appropriate RNA sequence is bound in their proximity. The change in fluorescence for residue F11W is intriguing as it is quenched by A_{15} and enhanced by U_6.

To ensure that the observed quenching is not simply an effect of nonspecific quenching by RNA nucleotides, N-acetyl-L-tryptophanamide (NAT) was used to mimic a Trp residue (Figure 20). A_{15} or U_6 was titrated into a 1 or 6 μM NAT sample (Figure 21). Neither RNA sequence induced quenching indicating that a specific or proximal interaction must occur to observe TFQ. Since significant quenching is only observed

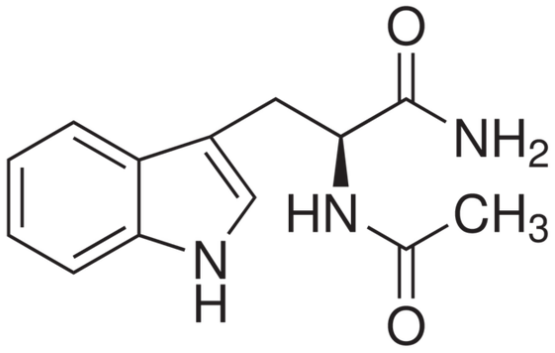


Figure 20: Structure of N-acetyl-L-tryptophanamide

The chemical structure of N-acetyl-L-tryptophanamide (NAT) mimics the structure Trp takes when it is connected to other amino acids to form a protein.

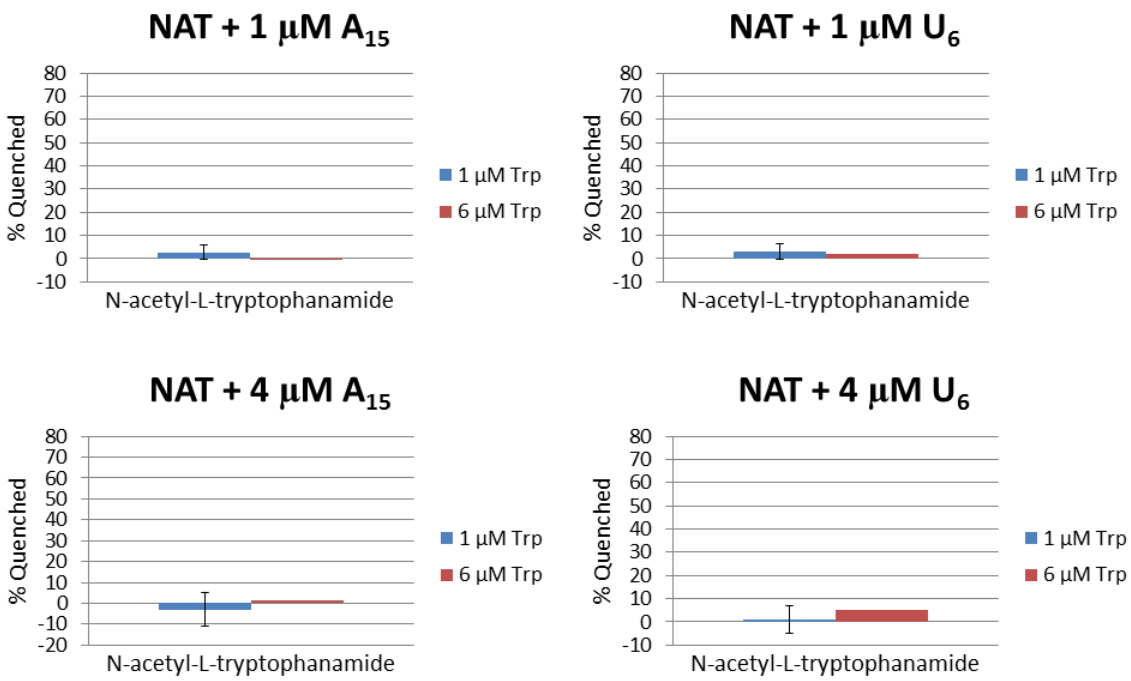


Figure 21: N-Acetyl-L-Tryptophanamide Quenching by A_{15} and U_6

TFQ experiments using NAT. A_{15} and U_6 were titrated into a 1 or 6 µM sample of NAT and examined for TFQ. The graphs illustrate no quenching for either substrate indicating that specific interactions between Trp and a nucleotide must occur to observe quenching.

for residues that are known from crystal structures to interact directly with RNA, we conclude that TFQ can be used to map accurately, albeit at lower resolution, specific RNA binding sites on Hfq.

2.3.2 *Staphylococcus aureus* Hfq

The main difference that exists between Gram-positive and Gram-negative homologues of Hfq resides in the distal face binding motif. In Gram-negative Hfq

homologues, specifically *E. coli*, the distal face has an $(A\text{-}R\text{-}N)_n$ RNA binding motif. In Gram-positive Hfq the sequence differences lead to a loss of the A site, and equivalent but not identical R-sites with the base stacking mode modestly different. In Gram-negative Hfq the main base-stacking residue in the R-site is L32 while in Gram-positive Hfq the main R-site base stacking residue is Y25 (Figures 7 & 8). Also in Gram-negative bacteria K31 does not interact with bound RNA while in Gram-positive bacteria Q31 stacks with the adenine ring of the adenosine in the L-site. This difference in the distal face binding mode should be possible to detect by TFQ since fluorescence quenching requires a direct interaction.

The *S. aureus* Hfq-A_4 structure indicates that residues F25 and Q31 stack with the R-site and L-site, respectively, and therefore they were mutated to F25W and Q31W for use in our TFQ assay *(62)*. Analysis of the *S. aureus* Hfq-AU_5G structure indicates that residue Y42 should be mutated (Y42W) to probe proximal face interaction *(38)*. Due to an inability to make *S. aureus* Y42W a second Gram-positive Hfq homologue from *L. monocytogenes* was used. Alignment of these two homologues reveals a 40% sequence identity and 81% sequence similarity. It also reveals that *L. monocytogenes* Hfq residue Y43 is equivalent to *S. aureus* residue Y42 (Figure 22). Thus *L. monocytogenes* Hfq F43W was used to probe proximal face binding by selected RNA sequences. The *S. aureus* Y25W, Q31W, and *L. monocytogenes* Y43W mutations were made in full length *S. aureus*

and *L. monocytogenes* Hfq, which are both 25 residues shorter than *E. coli* Hfq.

Figure 22: Sequence Alignment of *S. aureus* and *L. monocytogenes* Hfq Homologues

Sequence alignment was generated in CLC Sequence Viewer (CLC Bio, Cambridge, MA). The bottom row indicates those residues that are conserved between the two Hfq homologues. The red square highlights *S. aureus* Hfq residue Y42 and the corresponding residue Y43 in *L. monocytogenes* Hfq.

These mutants were then used in TFQ experiments with the control RNA sequences, A_{15} and U_6. These results indicate that it is indeed possible to properly identify the RNA binding surface by TFQ as both *S. aureus* Y25W and Q31W are significantly quenched in the presence of A_{15} and not by U_6 while *L. monocytogenes* Y43W is significantly quenched in the presence of U_6 and not A_{15} (Figure 23). These results also show that it is possible to use TFQ to discern the differences between the different distal face binding motifs for Gram-negative versus Gram-positive Hfq homologues. This is shown by the simple fact that TFQ detects significant quenching for *S. aureus* Q31W, which base stacks with RNA in the germane crystal structures, while TFQ detects no quenching for *E. coli* K31W, which is known to be physically close, ~ 6 Å, but not to interact with RNA in the germane crystal structures (Figures 19 & 23). This observation illustrates the strength of TFQ and supports the conclusion that these designed Trp mutants do not alter the known RNA binding motifs.

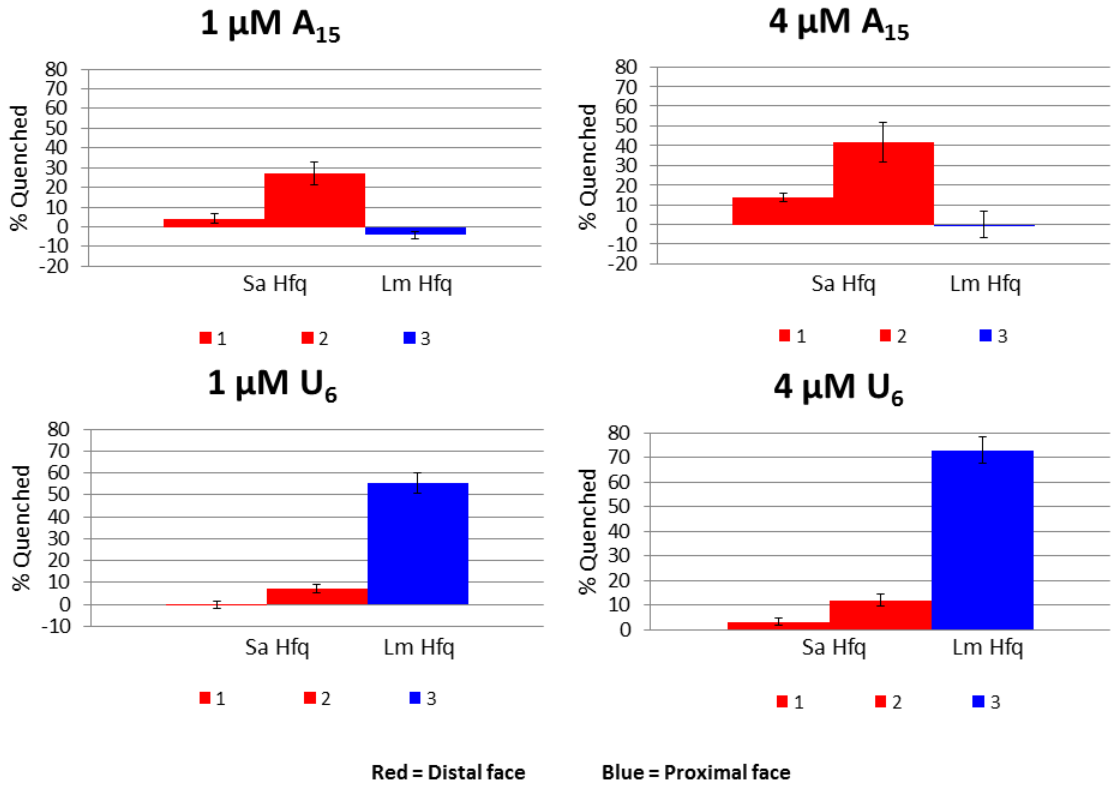


Figure 23: TFQ of Gram-positive Hfq homologues using A_{15} and U_6

Control TFQ experiments using the point mutants 1 = *S. aureus* F25W, 2 = *S. aureus* Q31W and 3 = *L. monocytogenes* F43W that were made in the full length *S. aureus* and *L. monocytogenes* Hfq homologues. The sequences A_{15} and U_6 were titrated from 1 to 4 µM. As these graphs illustrate, both sequences were observed to bind to the correct face.

2.4 The tryptophan mutants can reduce binding affinity but do not create false binding sites

Fluorescence Polarization (FP) is a technique that uses polarized light to excite a fluorophore which will then emit polarized light. If the fluorescent molecule is capable

of moving freely before emitting light then the polarization will be reduced *(63)*. The ability of a molecule to tumble freely in solution is dependent on its size and mass, a larger or less compact macromolecule will move more slowly than a small macromolecule *(60)*. Thus it is possible to measure the fluorescence polarization of a small substrate, titrate in a binding partner that is a larger size and observe the increase in polarization that happens. Polarization is defined by the equation: $P = (I_{\parallel}-I_{\perp})/(I_{\parallel}+I_{\perp})$ where P is polarization, $I_{\parallel}$ and $I_{\perp}$ are the intensities of the emission viewed through parallel and perpendicularly arranged polarizers, respectively. By titrating in a macromolecule of interest until polarization has stopped increasing a binding constant can be determined. To determine the binding constant of Hfq to target RNA sequences the RNA substrate is labeled with the 5-carboxyfluorescein (5-FAM) fluorophore and the polarization of RNA is monitored as Hfq is titrated until polarization stops increasing. The ensuing graph is then fit using the equation discussed in section 2.2.

To ensure that the designed Trp mutants did not significantly impact the function of Hfq, FP-based RNA binding assays were used to determine the equilibrium dissociation constants (K_d) of each mutant for the control RNA sequences A_{15} and U_6. Calculated affinities and the fold difference from WT binding affinities are shown in Table 2.

Table 2: Dissociation Constants (K_d) for Trp Mutants to rA_{15} and rU_6

	A_{15}		U_6	
	***E. coli* Hfq**			
WT Hfq S	1.4 nM ± 0.9[a]		766.0 nM ± 73.7	
F11W	7.1 nM ± 0.6	5.0	502.7 nM ± 55.5	0.7
R17W	3.7 nM ± 1.5	2.6	1.6 µM ± 0.08	2.1
Y25W	258.2 nM ± 32.7	184.4	836.8 nM ± 62.5	1.1
K31W	1.4 nM ± 0.5	1.0	1123.3 ± 73.0	1.5
Q33W	15.9 nM ± 3.6	11.4	404.7 nM ± 182.4	0.5
F39W	0.59 nM ± 0.35	0.4	2.0 µM ± 0.4	2.6
F42W	0.28 nM ± 0.09	0.2	4.8 µM ± 782.5	6.2
G77W	2.9 nM ± 1.7	2.1	379.9 nM ± 93.2	0.5
Y83W	7.4 nM ± 2.8	5.3	777.9 nM ± 254.1	1.0
Q95W	2.6 nM ± 1.1	1.9	848.3 nM ± 289.4	1.1
	***S. aureus* Hfq**			
WT Hfq	4.2 nM ± 0.5[b]		69.8 nM ± 7.0[b]	
F25W	11.3 µM ± 5.0	2700	41.9 nM ± 25.1	0.6
Q31W	1.6 µM ± 0.08	386.7	114 nM ± 23.2	1.6

[a] value taken from *(44)* [b] values taken from *(46)* Red highlights those residues that are displaying significant changes in K_d.

It was observed that mutating residue *E. coli* Y25W reduces the binding affinity of A_{15} 184-fold (from 1.4 nM to 258.2 nM) and mutating residue *E. coli* Q33W reduces the binding affinity of A_{15} 12-fold (from 1.4 nM to 15.9 nM) (Table 2). Mutating *S. aureus* F25W reduces A_{15} binding by 2,690-fold (from 4.2 nM to 11.3 µM) and mutating *S. aureus* Q31W reduces A_{15} binding by 381-fold (from 4.2 nM to 1.6 µM) (Table 2). The other

mutations do not adversely affect A_{15} binding. It was also observed that the *E. coli* F42W mutation reduces U_6 binding affinity 6-fold (766.0 nM to 4.8 μM), whilst the other residues show no adverse effects. We conclude that mutations involved in RNA binding show a decrease in binding affinity to the RNA sequences they bind. However, since the protein-RNA interaction is still observed via TFQ and florescence polarization it is concluded that this reduction in binding affinity is not detrimental to the ability of TFQ to detect which face a particular RNA sequence binds. It is noted that there is the chance for false negatives to occur, which indicates the importance of screening around the identified Hfq binding sites for the Trp mutants that least affect binding affinity while still detecting binding by TFQ. Additionally, it was also noted that mutations proximal to a particular binding site do not affect the affinity of distant sites for their preferred RNA sequence. For example, the *E. coli* F39W protein binds A_{15} with a K_d = 0.6 nM and the *E. coli* Y25W protein binds U_6 with a K_d = 836 nM, the same values as observed for wild type Hfq binding to these oligoribonucleotides.

2.5 Crystal Structures of Trp Mutants Show Minor Structural Perturbation

To verify that Trp mutations of *E. coli* Hfq do not significantly impact the overall global structure of the resulting Hfq protein, we determined the crystal structures of four key *E. coli* Trp mutants (F11W, Y25W, F39W, F42W) using the C-terminally truncated construct (residues 2-69) (PDB IDs: 4JRK, 4JUV, 4JRI, 4JLI, respectively).

These structures were then compared to the WT structure determined by Sauter *et al (39)* (PDB ID: 1HK9). All Root Mean Square Deviation (RMSD) for protomer:protomer and hexamer:hexamer for each Trp mutant as compared to WT can be found in Table 3.

Table 3: RMSD's of WT:Mutant *E. coli* Hfq Structures

	WT:Mutant RMSD	
	Monomer:Monomer	Hexamer:Hexamer
F11W	0.30 – 1.05 Å	2.48 Å
Y25W	0.27 – 0.64 Å	0.72 Å
F39W	0.37 – 0.54 Å	1.00 Å
F42W	0.31 – 0.52 Å	1.20 Å

E. coli Hfq Y25W crystallized in space group $P2_1$ and diffracted to 2.20 Å resolution. The superposition of Y25W and WT *E. coli* Hfq is shown in Figure 24A and indicates no structural changes. Alignment of the Y25W mutant with WT *E. coli* Hfq-A_{15} provides a rationale for the significant change we observe in the K_d of A_{15} for the mutant. The orientation of the Trp side chain is such that a shift in its conformation has to occur in order to avoid a clash with either the adenosine base or the 2' hydroxyl group of ribose (Figure 24B). *E. coli* Hfq F42W crystallized in the rhombohedral space group R3 and diffracted to 1.79 Å resolution. Superposition of F42W and WT *E. coli* Hfq is shown in Figure 24A and indicates no structural changes. This alignment also indicates that there would be little change in the binding and recognition of uridine, supporting the slightly lower binding affinity. The F39W mutant crystallized in the cubic space group

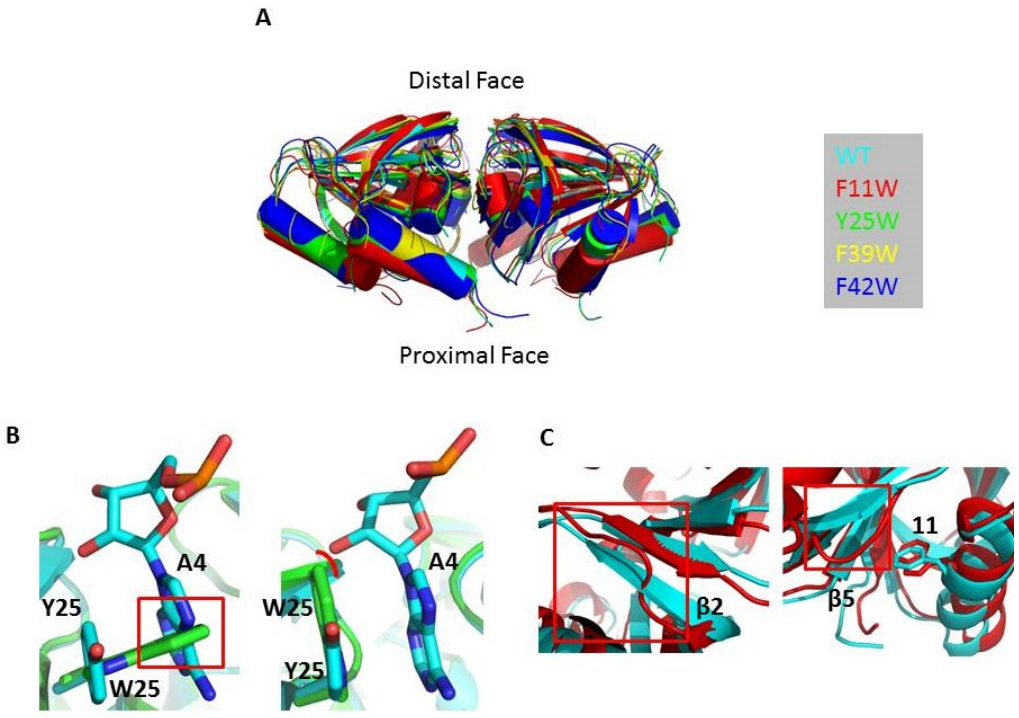


Figure 24: Overlay of Trp Mutant Structures to WT *E. coli* Structure

(**A**) Overlay of the hexamer structures of WT *E. coli* Hfq and the selected mutants: F11W, Y25W, F39W, and F42W. All proteins are shown as cartoons with WT *E. coli* Hfq colored cyan, Y25W colored green, F11W colored red, F39W colored yellow, and F42W colored blue. The calculated RMSDs reveal that each mutation does not affect the protomeric structures. (**B**) Two close up views of the overlay of the structures of the WT *E. coli* Hfq bound to A_{15} and the Y25W mutation near the position of the substitution and R-site. Here it can be seen that W25 takes two conformations with one monomer having both conformations present. One confirmation occurs in 2 out of 6 protomers (left) and would block adenine insertion into the R-site. The second conformation (right) occurs in 5 out of 6 protomers and allows stacking with adenine. However the 2' ribose oxygen clashes with the indole ring, requiring some phosphodiester or polypeptide backbone adjustment to relieve the clash. These two observations are likely the cause for the significant reduction in A_{15} binding to the distal face mutant. (**C**) Close up of the overlay of the apo WT *E. coli* Hfq and F11W hexamers about β2 and β5. Highlighted by red boxes are the two major conformational differences between the F11W protein and WT *E. coli* Hfq.

$P2_13$ and diffracted to 1.86 Å resolution. Superposition of F39W and WT *E. coli* Hfq is shown in Figure 24A and indicates reasonable structural identity. The F11W mutant crystallized in the orthorhombic space group $P2_12_12$ and diffracted to 1.93 Å resolution. When individual subunits are aligned with those of WT *E. coli* Hfq the RMSD is 0.30-1.05 Å. Interestingly, alignment of the F11W hexamer and the WT *E. coli* Hfq hexamer results in an RMSD of 2.48 Å indicating a significant quaternary change (Figure 24A). The structure of F11W provides the reason for this significant difference, although it cannot be ruled out that these structural changes may be due to crystal packing effects. Substitution of a Trp residue at position 11 causes a significant shortening of two β strands by the adjustment of the backbone about residues Q35 to F39 in β strand β2 (Figure 24C) and residues I59 to V62 in the β strand β5 (Figure 24C). These strands form part of the distal-side intersubunit interface, which is now partially disturbed. Some change might have been expected from modeling studies given the bulkier nature of Trp but underscores the importance of having the high-resolution structure of this particular mutant. Regardless, it is likely the remodeling of the β2 and β5 strands and their interaction with adjacent subunits are the origin of the perturbed, but not titratable, F11W TFQ data.

2.6. Conclusions and Future Directions

The majority of wild type Hfqs have no native Trp. This makes it straight forward to design Trp mutants, based on known structures and sequence alignments, and to use those trp mutants in order to map Hfq-RNA binding sites using TFQ. TFQ will only be observed if a direct interaction occurs between the Trp residue and RNA as was shown by using NAT in control TFQ experiments (Figure 21). Several different point mutants were created in the full length *E. coli*, *S. aureus*, and *L. monocytogenes* Hfq proteins. These mutants were designed to cover the distal and proximal binding faces, a proposed lateral binding site in *E. coli* Hfq and the relatively long C-terminus in *E. coli* Hfq in order to explore if and how different RNA sequences may interact with these surfaces. Using the control sequences A_{15} and U_6, which bind the distal and proximal faces respectively; these mutants were examined for quenching. Here it was shown that the TFQ experiment correctly identified where a particular RNA sequence will bind. It was also observed that the TFQ experiments can differentiate between the two different known distal face binding motifs for Gram-negative (*E. coli)* and Gram-positive (*S. aureus* and *L. monocytogenes*) as we detected that residue 31 does not experience quenching for the *E. coli* Hfq while it sees significant quenching for *S. aureus* Hfq when A_{15} is titrated in. The germane crystal structures show that residue 31 does not interact

with RNA in the *E. coli*-A_{15} structure while it is directly contacted by RNA in the *S. aureus* Hfq-A_4 structure (Figure 25).

To ensure that the Trp mutations did not destroy function, binding affinities for all Trp mutants to the control sequences were calculated (Table 2). It was observed that those mutants responsible for binding do experience a change in binding that could result in lost binding for some weakly interacting RNA ligands. Overall however the

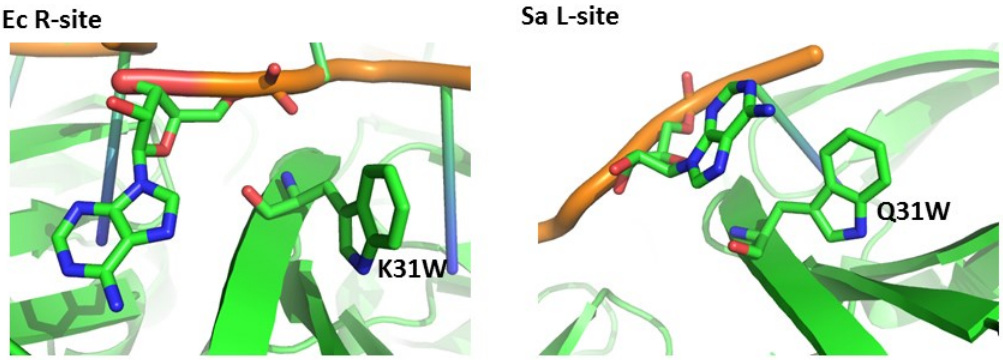


Figure 25: Comparison of the locations of residue 31 in the *E. coli* Hfq – A_{15} and *S. aureus* Hfq – A_4 structures

Residues *E. coli* K31 and *S. aureus* Q31 were modeled as Trp point mutants to analyze the differences in how these residues might interact with the bound ribonucleotide. It is shown that *E. coli* K31W cannot stack with the bound adenosine while *S. aureus* Q31W can stack the bound adenosine thus suggesting that *S. aureus* Q31W will quench in the presence of adenosine and *E. coli* K31W will not. This is in agreement with the observations from TFQ experiments. These models were generated in Pymol using PDB IDs: 3GIB (*E. coli* Hfq-A_{15}) *(44)* and 3QSU (*S. aureus* Hfq-A_4) *(46)*.

vast majority of mutants do not impact function and RNA binding is observed for all mutants to all proper sequences. To further verify that these mutants have not

disrupted Hfq function the structures of selected mutants were determined and, with the exception of F11W, only slight variations in the hexameric structures were seen. Taken together all of this data indicates that TFQ correctly identifies the face that a particular RNA sequence will bind to on Hfq in a simple, rapid manner.

Future plans include expanding the mutants available for analyzing RNA binding to the *S. aureus* and *L. monocytogenes* Hfq homologues. Further, it would be ideal to find a distal face mutant that will identify binding but not significantly impact RNA binding affinity. Determining the apo crystal structures of all Trp mutants should be done to ensure that none of the mutants are significantly affecting the structure. It is also of interest to determine the RNA bound crystal structure for those mutants that are involved in binding, for example determining the *E. coli* Q33W-A_{15} structure to see if the structure is changing upon binding. Finally, as this thesis will expand upon in later chapters, it will be most interesting to use TFQ to study further the proposed distal face binding motifs and to analyze how physiologically relevant larger RNA sequences interact with Hfq.

3. The Hfq Distal Face Binding Motif

3.1 Introduction to the Two Major Distal Face Binding Motifs

The current structures for Hfq homologues bound to RNA sequences on the distal face exhibit two different binding motifs that are related to whether the Hfq homologue comes from a Gram-negative or Gram-positive bacterium (Figure 26). In the Gram-negative structure of *E. coli* Hfq-A_{15} the observed binding on the distal face is an $(A\text{-}R\text{-}N)_n$ motif where A=adenine nucleotide, R=any purine nucleotide and N= any nucleotide *(44)*. This motif indicates that there are a maximum of 18 binding sites on the distal face of the hexamer.

The A-site specificity of *E. coli* Hfq is conferred by peptide backbone amide and carbonyl oxygen hydrogen bonding between residue Q33 and the N7 and N6 atoms of adenine and a polar interaction between the Nε amide of residue Q52 and the adenine N1 atom (Figure 27). The R-site of *E. coli* Hfq appears to be able to accommodate both adenosine and guanosine whereby the purine ring sticks into a pocket and packs against residues Y25, L26', I30' and L32', where the prime indicates residues from the neighboring subunit. The adenosine N3, N6, N1 and ribosyl 2' hydroxyl atoms Hydrogen bond to Nδ of residue N28, Oε atom of residue Q52, Oγ of residue T61, and the carbonyl oxygen of residue G29, respectively (Figure 27). The N site has no protein-nucleic acid interactions and likely represents the entrance or exit point for RNA.

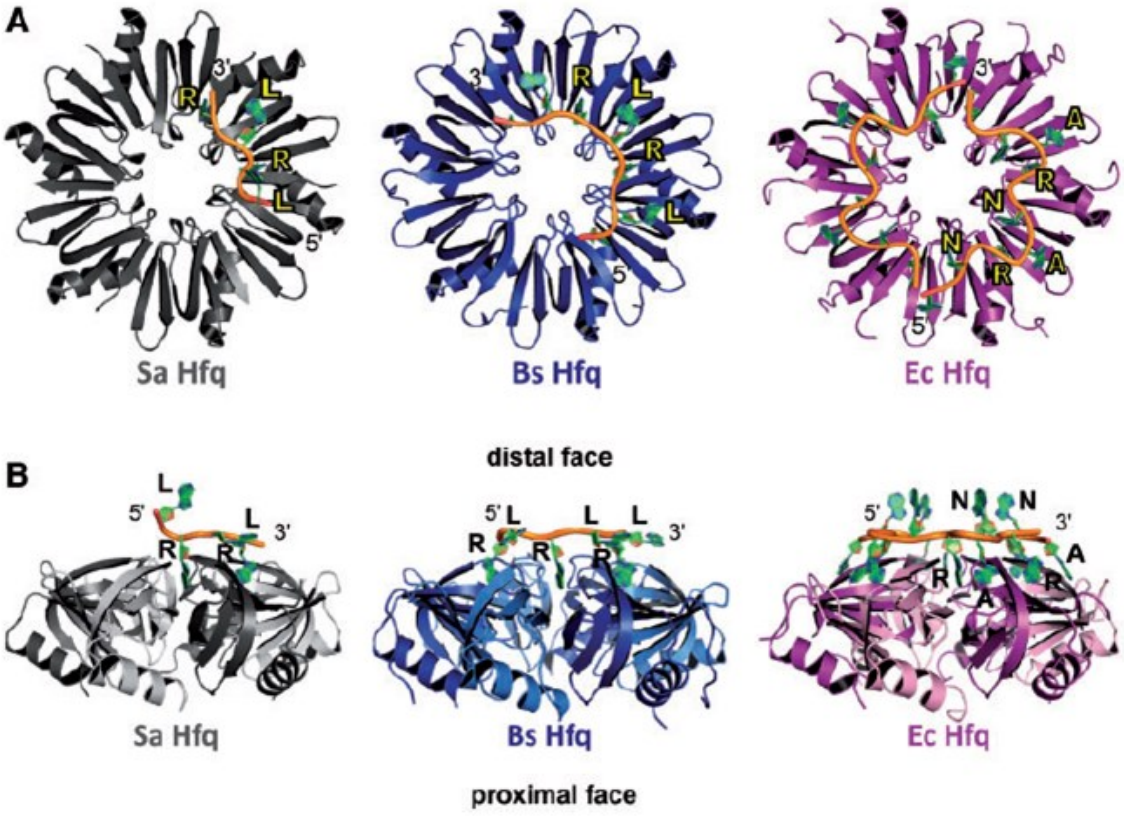


Figure 26: Gram-positive and Gram-negative Distal Face Binding Motifs

The RNA distal face binding motifs of Hfq in Gram-positive and Gram-negative bacteria. (**A**) View looking at the distal face of *S. aureus* Hfq (left, grey), *B. subtilis* Hfq (middle, blue), and *E. coli* Hfq (right, magenta). Bound RNA is shown as a cartoon with the phosphate backbone colored yellow and the purine bases, shown as sticks, colored green. A= adenine nucleotide only, R = any purine nucleotide, L = linker nucleotide, N = any nucleotide. The 5′ and 3′ ends of the RNA strand are labeled. (**B**) Side view of the *S. aureus* Hfq-A_4 (left, grey), *B. subtilis* Hfq-$(AG)_3A$ (middle, blue), and *E. coli* Hfq-A_{15} (right, magenta) complexes. Each Hfq is labeled as in A and the distal and proximal faces are labeled. Taken from *(46)*

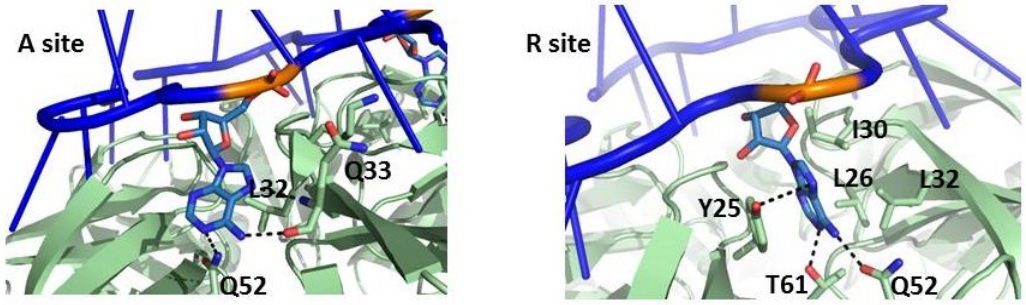


Figure 27: *E. coli* Hfq A-site and R-site interactions

The adenine nucleotide specificity (A-site) and purine nucleotide selectivity (R-site) sites of *E. coli* Hfq. In both figures Hfq is shown as a cartoon with the residues used for selectivity and adenosine shown as sticks. Hfq is colored light green while the RNA is colored blue with the atoms for the side chains colored by atom type (carbon is green for Hfq and blue for RNA, oxygen is red, phosphorus is orange). Dotted lines represent Hydrogen bonds. Images were made in Pymol using PDB ID: 3GIB *(44)*.

In the two structures of gram-positive Hfqs, *S. aureus* Hfq-$(AA)_3A$ and *B. subtilis* Hfq-$(AG)_3A$, the observed binding motif has been shown to be an $(R\text{-}L)_n$ motif where R is any purine nucleotide and L is a linker nucleotide (Figure 26). The R-site is located between β-strands 2 and 2′ from neighboring subunits. The bound adenosine is found to base stack with residues F25, F26′ and F30′, where the prime indicates residues from the neighboring subunit, and make small van der Waals contacts with L27′ and M32′ (Figure 28). Hydrogen bonds for the R-site are observed between the side chain amide of N28′, the sugar O4′ and adenine N3 atoms and between the 2′-OH group and G29 (Figure 28). Finally there are hydrogen bonds between the N1 and exocyclic N6 amino group of the adenine and the hydroxyl groups of S61 and T62, respectively (Figure 28).

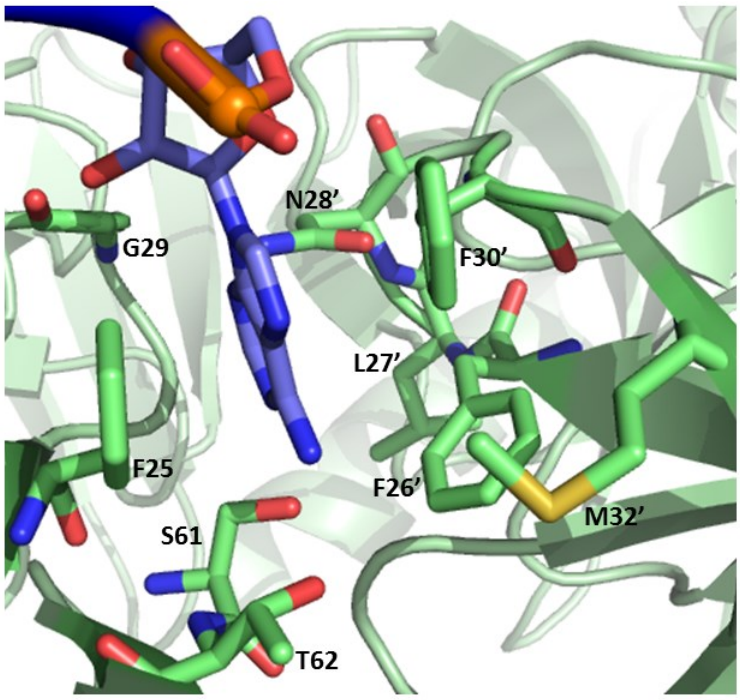


Figure 28: *S. aureus* Hfq R-site Interactions

The purine selectivity site of *S. aureus* Hfq. Hfq and RNA are shown as a cartoon with the selective residues and adenosine shown as sticks. Hfq is colored green while RNA is colored blue. The sticks are colored such that carbon is green for Hfq and blue for RNA, oxygen is red, nitrogen is blue, phosphorus is orange and sulfur is yellow. Image was made in Pymol using PDB ID: 3QSU *(46)*.

All of these interactions between RNA and the R-site of Hfq anchor the RNA to the distal face of Hfq. For both distal face binding motifs the R-site is capable of accommodating a guanosine, although some movement must occur in the *S. aureus* Hfq-A_4 structure (Figure 29). However, the currently available structures all have adenosine

bound in the R-site. Currently the only data that supports the proposed distal face (A-R-N)$_n$ and (R-L)$_n$ motifs are RNA binding assays that determine binding affinity but do not confirm where the RNA sequence is actually binding on Hfq. Thus since currently it is not known whether guanosine will actually bind in the R-site of the distal face motifs our TFQ method was applied to study how RNA sequences that fulfill these motifs interact with Hfq.

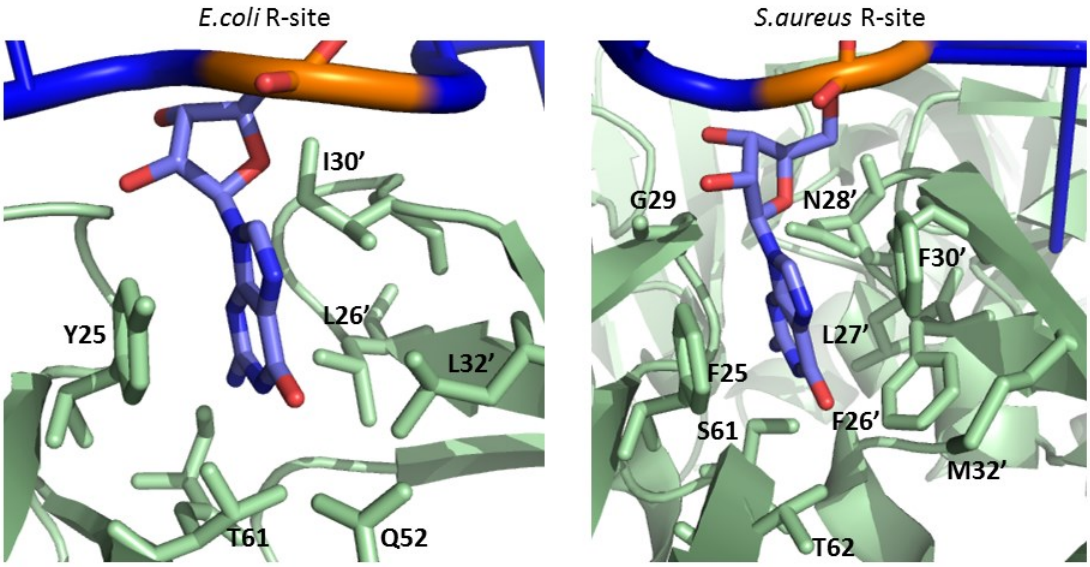


Figure 29: *E. coli* and *S. aureus* Hfq with Modeled Guanosine in R-site

The purine selectivity (R-site) sites for both *E. coli* and *S. aureus* Hfq with guanosine modeled into the site. In both figures Hfq and RNA are shown as cartoons with the selective residues shown as sticks. Hfq is colored light green while the RNA is colored blue. Adenosine is colored by atom such that carbon is light blue, oxygen is red, nitrogen is blue and phosphorus is orange.

3.2 E. coli Hfq Distal Face Binding is Restricted to (A-A-N)$_n$ Motifs

To test the proposed (A-R-N)$_n$ distal face motif for *E. coli* Hfq two different RNA sequences were used, 5′ - AAGAAGAAGAAGAAG - 3′ (A-A-G)$_5$ and 5′ - GGAGGAGGAGGAGGA - 3′ (G-G-A)$_5$. The (A-A-G)$_5$ motif fulfills both the (A-R-N)$_n$ motif and an (A-A-N)$_n$ motif while the (G-G-A)$_5$ sequence only fulfills the (A-R-N)$_n$ motif. As expected the (A-A-G)$_5$ sequence was observed to quench residues F11W and Q33W but not residues F39W or F42W indicating that (A-A-G)$_5$ binds to the proximal face (Figure 30). Somewhat surprising was that (A-A-G)$_5$ does not quench residue Y25W, which also lies on the distal face. However, given that mutating Y25W causes a 180-fold

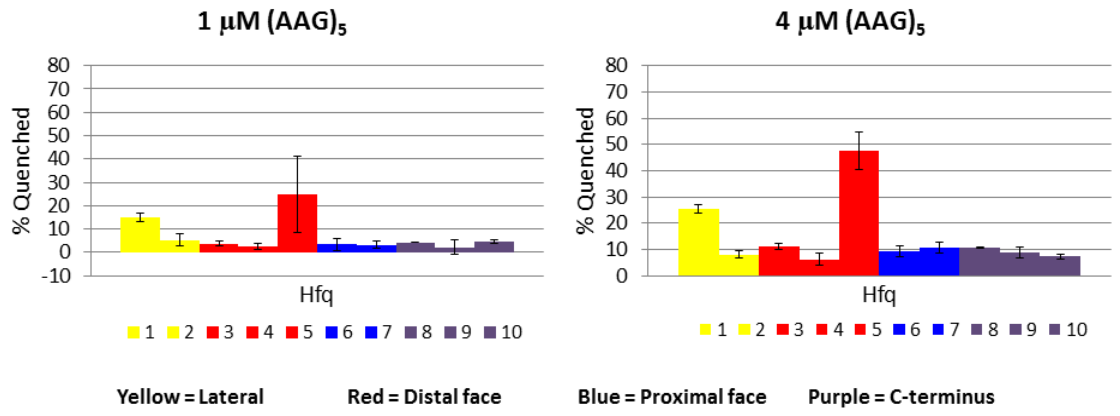


Figure 30: *E. coli* Hfq TFQ with (A-A-G)$_5$ RNA

TFQ of *E. coli* Hfq Trp point mutants using the sequence (A-A-G)$_5$. 1 = F11W, 2 = R17W, 3 = Y25W, 4 = K31W, 5 = Q33W, 6 = F39W, 7 = F42W, 8 = G77W, 9 = Y83W, 10 = Q95W. Here it is observed that (A-A-G)$_5$ quenches F11W and Q33W mutants indicating that (A-A-G)$_5$ binds on the distal face of *E. coli* Hfq as expected.

reduction in binding affinity of A_{15}, RNA sequences with weaker binding affinity, such as $(A\text{-}A\text{-}G)_5$, are likely to be unable to bind well even at 4 μM and hence not bind nor quench this R-site substitution. Regardless distal face binding is supported by the strong quenching of distal face residue Q33W (Figure 30).

In contrast to the observed distal face quenching for $(A\text{-}A\text{-}G)_5$ when $(G\text{-}G\text{-}A)_5$ is used for TFQ strong quenching is observed for residue F42W and not for Y25W or Q33W supporting proximal face binding only (Figure 31). $(G\text{-}G\text{-}A)_5$ is also observed to quench residue Y83W in the C-terminal tail suggesting that longer RNA sequences that bind the proximal face may interact with the C-terminal tail, or that they bind the

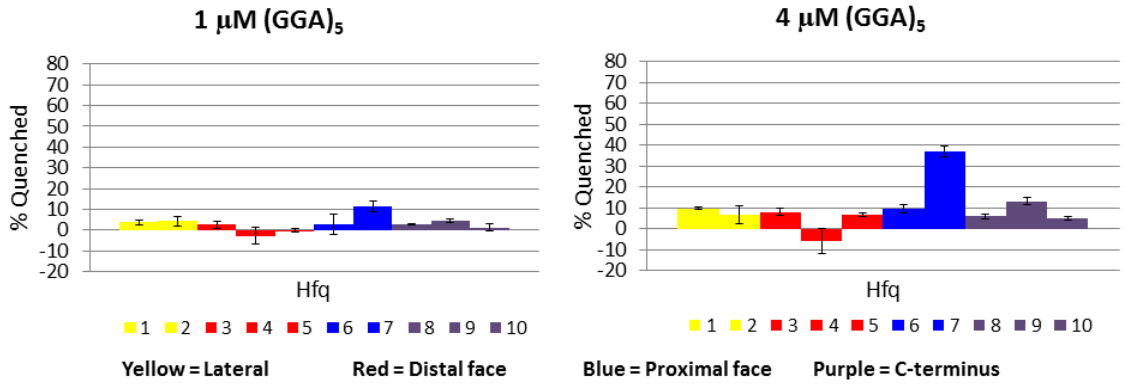


Figure 31: *E. coli* Hfq TFQ with $(GGA)_5$ RNA

TFQ of *E. coli* Hfq Trp point mutants using the sequence $(G\text{-}G\text{-}A)_5$. 1 = F11W, 2 = R17W, 3 = Y25W, 4 = K31W, 5 = Q33W, 6 = F39W, 7 = F42W, 8 = G77W, 9 = Y83W, 10 = Q95W. It is observed that $(G\text{-}G\text{-}A)_5$ quenches F42W and Y83W mutants indicating that $(G\text{-}G\text{-}A)_5$ binds on the proximal face of *E. coli* Hfq and interacts with part of the C-terminal tail.

C-terminus independently but weakly. Interestingly residues G77W, Y83W and Q95W are not quenched in the presence of $(A\text{-}A\text{-}G)_5$ but residue G77W is quenched in the presence of $(A\text{-}A\text{-}A)_5$ suggesting that these two distal face binding sequences may utilize different entrance/exit pathways or that an adenine but not guanosine is recognized by some residues near G77.

From these data we can conclude that the *E. coli* Hfq distal face binding motif is a more restrictive $(A\text{-}A\text{-}N)_n$ motif rather than an $(A\text{-}R\text{-}N)_n$ motif and that the R-site is an adenosine only binding pocket. These results also demonstrate that some C-terminal residues are involved in the interaction with $(A\text{-}A\text{-}A)_n$ and $(G\text{-}G\text{-}A)_n$ tracts implying the C-terminal region of Hfq can interact with longer proximal face binding RNA.

3.3 Proposed S. aureus Hfq Distal Face $(R\text{-}L)_n$ Motif is Restricted to an $(A\text{-}L)_n$ Motif

To test the proposed $(R\text{-}L)_n$ binding motif of Gram-positive bacterial Hfq the sequences 5′ – AUAUAUA – 3′ $(A\text{-}U)_3A$, 5′ – ACACACA – 3′ $(A\text{-}C)_3A$, 5′ – GGGGGGG – 3′ $(G\text{-}G)_3G$ and 5′ – GUGUGUG – 3′ $(G\text{-}U)_3G$ were tested by TFQ. All of these sequences fulfill the proposed $(R\text{-}L)_n$ binding motif. The $(A\text{-}U)_3A$ and $(A\text{-}C)_3A$ sequences also fulfill an $(A\text{-}L)_n$ motif while $(G\text{-}G)_3G$ and $(G\text{-}U)_3G$ sequences do not. From our analysis of the *S. aureus* Hfq-A_4 and *B. subtilis* Hfq-$(A\text{-}G)_3A$ complex structures the more restrictive $(A\text{-}L)_n$ motif is suggested. In these structures each one has an adenosine stacking in the R-

site over residue F25W whilst the 3′ nucleotide, an A or G, is sitting directly over the L-site forming residue Q31W. The TFQ results indicate that $(A\text{-}U)_3A$ and $(A\text{-}C)_3A$ sequences quench the distal face residues *S. aureus* Y25W and Q31W while $(G\text{-}G)_3G$ and $(G\text{-}U)_3G$ do not quench either of these residues (Figure 32). $(A\text{-}U)_3A$ and $(G\text{-}U)_3G$ were observed to quench the proximal face residue *L. monocytogenes* F43W while $(A\text{-}C)_3A$ and $(G\text{-}G)_3G$ do not (Figure 32). From these observations it can be concluded that the distal face binding motif for *S. aureus* Hfq, and likely most other Gram-positive bacterial Hfqs, is a more restrictive $(A\text{-}L)_n$ motif. The fact that $(A\text{-}U)_3A$ quenches both faces is not a surprise as it has both adenosines and uridines and adenosine is capable of binding to either face, as seen in the *S. aureus* Hfq-AU_5G structure *(38)* and the $(A\text{-}U)_3A$ sequence fulfills the distal face binding motif. The observation that $(G\text{-}U)_3G$ binds to the proximal face of *L. monocytogenes* Hfq supports the hypothesis that the $(A\text{-}U)_3A$ sequence binds to the proximal face through the uridines and is not inhibited by the presence of adenosine or guanosine. Finally the observation that $(G\text{-}G)_3G$ does not bind to either face is consistent with binding data that indicate that G_7 does not bind to Hfq *(38)*. To understand the binding mechanism of $(G\text{-}U)_3G$ high-resolution structural studies will need to be performed.

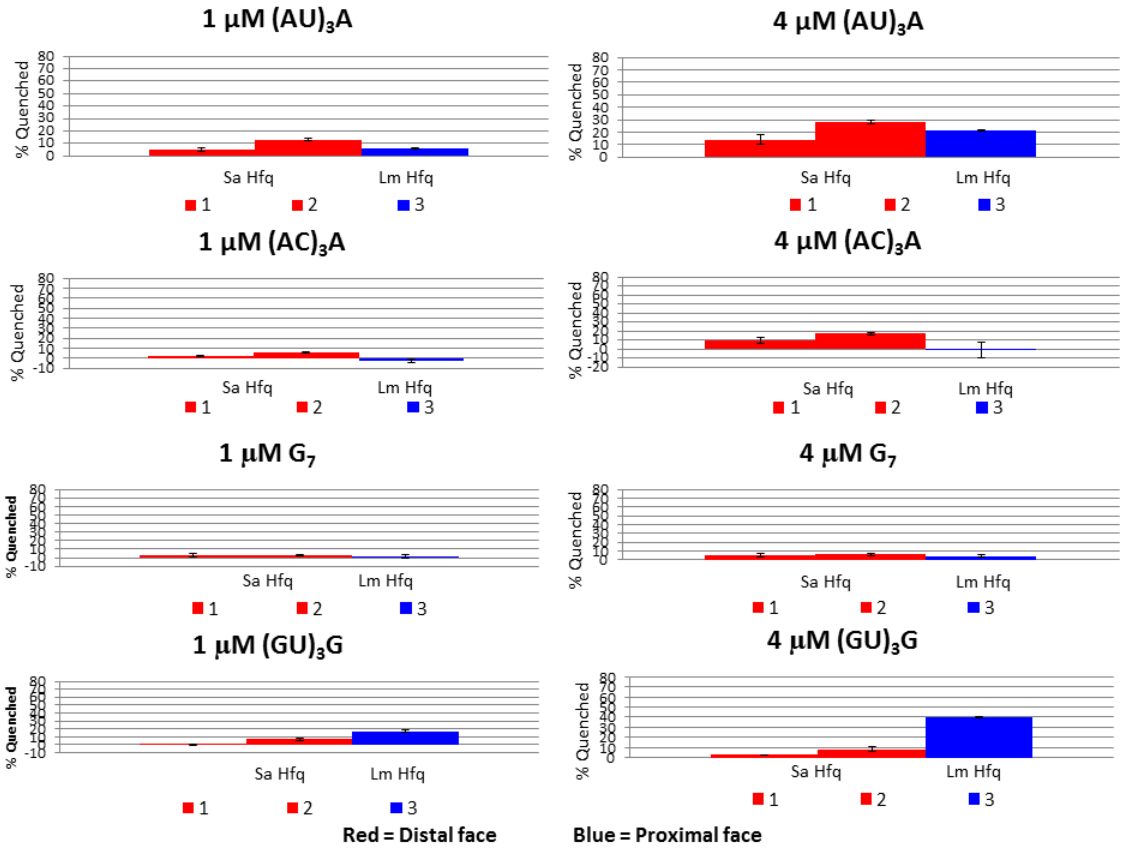


Figure 32: *S. aureus* Hfq TFQ Analysis of the Proposed (R-L)$_n$ binding motif

TFQ of *S. aureus* Hfq Trp point mutants using the sequences (A-U)$_3$A, (A-C)$_3$A, (G-G)$_3$G and (G-U)$_3$G. 1 = *S. aureus* F25W, 2 = *S. aureus* Q31W and 3 = *L. monocytogenes* F43W. It is observed that (A-U)$_3$A and (A-C)$_3$A quench the Y25W and Q31W mutants. (A-U)$_3$A and (G-U)$_3$G are observed to quench the proximal pore F43W while (G-G)$_3$G does not quench any residues.

3.4 Conclusions and Future Directions

Taken as a whole the data presented here illustrate that the proposed distal face binding motifs for both Gram-negative and Gram-positive bacterial homologues of Hfq can be redefined to a more restrictive sequence than previously hypothesized. For *E.*

coli, and likely most other Gram-negative bacterial Hfqs, the distal face motif is restricted to an $(A\text{-}A\text{-}N)_n$ motif from the previously proposed $(A\text{-}R\text{-}N)_n$ motif. For *S. aureus* and likely most other Gram-positive bacterial Hfqs the distal face motif is restricted to an $(A\text{-}L)_n$ motif from the previously hypothesized $(R\text{-}L)_n$ motif. To further verify that these proposed restricted binding motifs do indeed apply to other homologues of Hfq it is necessary to create the same mutants in at least one other homologue representing Gram-negative bacteria, such as *Salmonella typhimurium* (*S. typhimurium*), and one other homologue representing Gram-positive bacteria, such as *L. monocytogenes*.

The ability to refine the distal face motifs further demonstrates the strength of TFQ for examining RNA binding to Hfq. It has been previously hypothesized that the R-site was in reality an A-site due to the fact that all germane crystal structures have an adenosine bound in the R-site. However, since in these structures it is feasible to model a guanosine in the R-site it was not logical to restrict the R-site to adenosine. The distal face motifs have been extensively studied by binding assays and have illustrated that sequences which fulfill an $(R\text{-}L)_n$ motif do indeed bind to Hfq. These techniques however do not provide any information on where a particular sequence is binding Hfq, they just indicate if a sequence binds. TFQ is a simple, quick method for determining where a particular RNA sequence is binding.

The observation that the sequence $(G\text{-}G\text{-}A)_5$ binds to the proximal face rather than the distal face of *E. coli* Hfq is intriguing. The proximal face tends to bind uridine rich sequences and there is currently no high resolution structure of Hfq that has guanosine bound on the proximal face of a Gram-negative Hfq homologue. Thus understanding how the sequence $(G\text{-}G\text{-}A)_5$ is binding on the proximal face of *E. coli* Hfq warrants further investigation and necessitates high-resolution structures of this interaction. The observation that the sequence $(G\text{-}U)_3G$ binds to the proximal face of *L. monocytogenes* is also intriguing and will warrant further study as there are no high resolution structures of G-rich sequences bound to *L. monocytogenes* on the proximal face. The observation that both Gram-negative and Gram-positive Hfq homologues bind G-rich sequences on the proximal face is intriguing and may suggest that they have a similar binding mechanism.

These data further emphasize the utility of the TFQ technique as it further demonstrates that this technique is capable of differentiating between the two different binding motifs that exist for *E. coli* and *S. aureus*. This is best shown by the observation that in *E. coli* residue 31 is within 6 Å of the bound RNA but does not directly interact while for *S. aureus* residue 31 is the main residue involved in base stacking for the L-site and is quenched by cognate RNA (Figure 25). Here the tested sequences for *E. coli* do not interact with residue 31, thus indicating that the observed binding motif, while now

more refined, utilizes the same pathway. The tested sequences that bind to the distal face for *S. aureus* show strong binding to residue 31 indicating that for all distal face binding sequences this residue has a significant role in binding and these RNA sequences most likely use the same binding motif.

4. Autoregulation of *E. coli* Hfq

4.1 Introduction to Hfq Autoregulation

Hfq homologues have been shown to regulate the expression of many different mRNA transcripts. There are a variety of different mRNA transcripts that are regulated by a particular Hfq homologue however, not all homologues regulate the exact same transcripts. It has been shown that *E. coli* and *Sinorhizobium meliloti* (*S. meliloti*) are involved in autoregulation. Autoregulation means that the protein expressed from a particular mRNA transcript will feedback to control its own expression either raising, positive autoregulation, or lowering, negative autoregulation, its expression. For both *E. coli* and *S. meliloti* Hfq negative autoregulation is observed meaning that when Hfq is

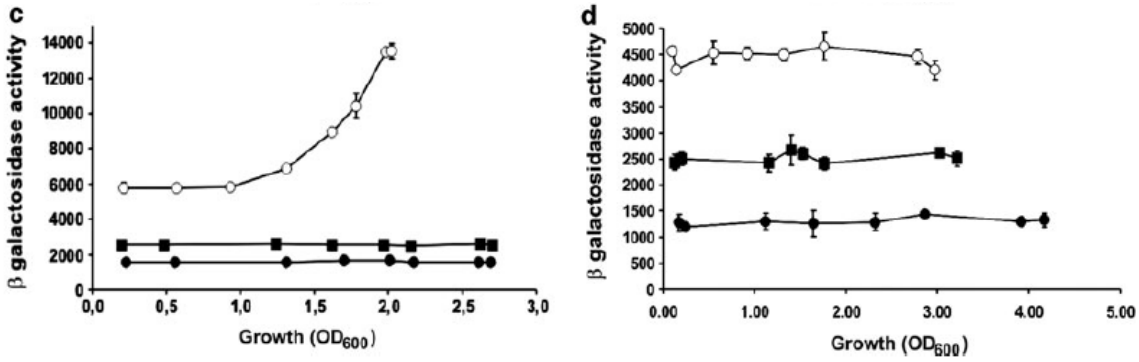


Figure 33: *S. meliloti* Hfq β-galactosidase assay illustrating negative autoregulation

β-galactosidase activity was assayed *in vivo* using *S. meliloti* strains 20PS15 (WT: filled circles), 20PS16 (*hfq* knockout strain: open circles), and 20PS16/p*hfq*$_{Sm}$ (complemented *hfq* knockout strain: filled square) that have been transformed to incorporate a sequence that uses the *hfq* promoter region to express the lacZ gene into the genome for performing a β-galactosidase assay. Image taken from *(7)*.

present it will bind to its own mRNA transcript and inhibit translation. For *S. meliloti* autoregulation has been observed *in vivo (7)*. The *hfq* promoter was used to control the *lacZ* gene. It was observed that a wild-type strain of *S. meliloti* will maintain a constant level of β-galactosidase activity while in a strain of *S. meliloti* that has *hfq* deleted, the β-galactosidase activity significantly increases during growth (Figure 33) *(7)*. It was also determined that adding *hfq* back to the deletion strain will return β-galactosidase

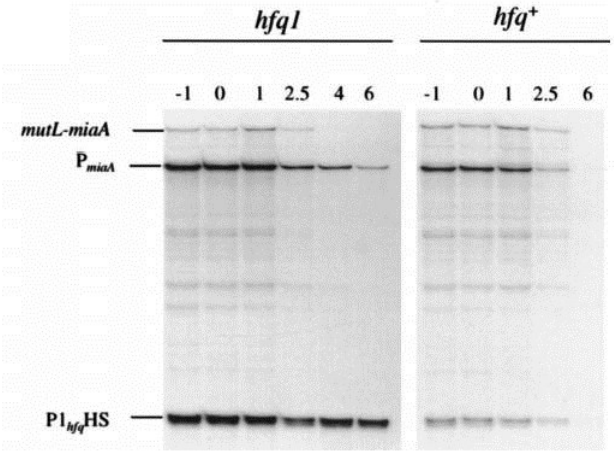


Figure 34: *In vivo* analysis of *hfq* mRNA transcript levels in *E. coli* Hfq null strain

Amount and stability of the $P1_{hfq}HS$, P_{miaA} and *mutL-miaA* transcripts in the *E. coli hfq1* knockout strain as detected by RNase T_2 protection assays. *Hfq1* is the Hfq null strain while *hfq*⁺ is the wild-type *E. coli* strain. Time points were taken just before (-1), just after (0) and 1, 2.5, 4 and 6 minutes after addition of rifampicin. Transcript levels for both $P1_{hfq}HS$ and P_{miaA} remain constant in the *hfq* knockout strain while they disappear over time in the wild-type strain indicating that the presence of *hfq* causes those transcripts to disappear. However the *mutL-miaA* transcript level does not change over time indicating that Hfq is not the cause for turnover of this transcript. Image taken from *(64)*.

activity levels to the constant wild type levels (Figure 33) *(7)*. Taken together these data indicate that *S. meliloti* Hfq negatively autoregulates itself.

For *E. coli* negative autoregulation was demonstrated *in vivo* by carefully inserting a kanamycin cassette into the middle of the *hfq* gene within the genome and monitoring *hfq* transcript levels present during growth *(64)*. In wild type *E. coli* cells there is an initial detection of *hfq* and over time the transcript disappears, whereas in the Hfq null mutant the transcript level remains constant (Figure 34). There are two binding sites within the 5′ untranslated region (5′ UTR) of *E. coli hfq* that have been shown to be important to the ability of Hfq to bind its mRNA transcript termed site A and site B (Figure 35). These two binding sites were identified by gel shift mobility assays *(65)*.

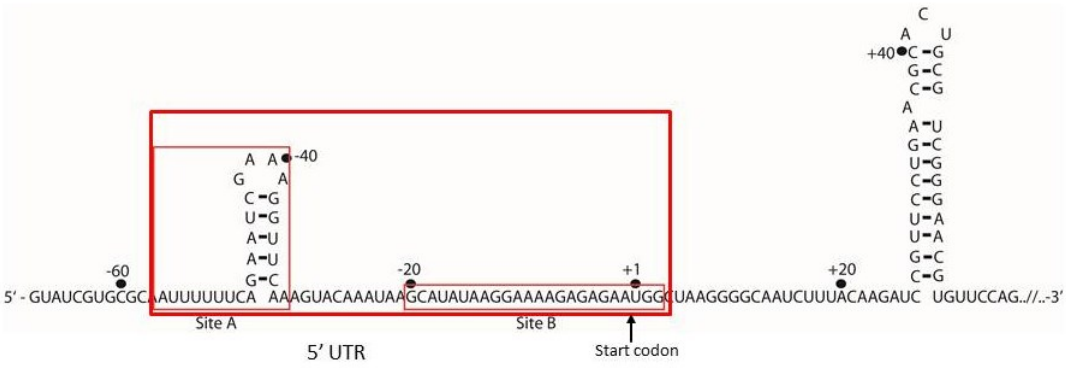


Figure 35: *E. coli* 5′ UTR of *hfq* mRNA transcript

The 5′ UTR sequence of *hfq* mRNA. The two identified Hfq binding sites, Site A and Site B, are highlighted with thin red boxes while the sequence termed 5′ UTR is highlighted with a thick red box. The start codon of the *hfq* gene is indicated by the black arrow. Image adapted from *(34)*.

Furthermore it has been determined *in vitro* that having Hfq in an *in vitro* translation initiation assay will prevent translation initiation, most likely by blocking the ribosome binding site (Figure 36) *(65)*. The observation that there are two binding sites would suggest that each site will bind to a particular face of Hfq. To date no high-resolution structure has been obtained that shows how these two binding sites interact with Hfq. Understanding how *E. coli* Hfq binds to its 5′ UTR, which is important for autoregulation, will improve our knowledge of how Hfq functions *in vivo*. By improving our knowledge of how Hfq autoregulates it may become possible to control Hfq autoregulation and to then potentially target Hfq levels as part of drug therapy treatment against multidrug resistant bacteria.

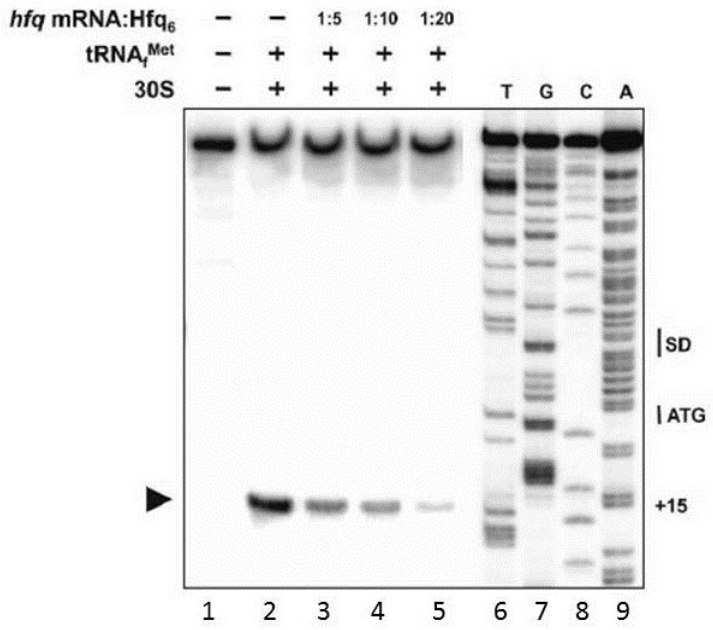


Figure 36: Translation initiation of *hfq* with and without hfq *(34)*

Toeprinting assay to illustrate autogenous inhibition of translation initiation by Hfq. Lane 1, primer extension in the absence of 30S subunits, $tRNA_f^{Met}$ and Hfq. Lane 2, toeprinting in the presence of 30S subunits, $tRNA_f^{Met}$, and absence of Hfq. Lanes 3–5, toeprinting with 30S subunits, $tRNA_f^{Met}$, and increasing molar Hfq-hexamer ratios to *hfq* mRNA as indicated at the top of the autoradiograph. Hfq was added prior to the addition of ribosomes. On the left the arrowhead depicts the toeprint signal at position +15 relative to the A of the ATG start codon. The relevant part of the DNA sequence of the *hfq* 5′ UTR along with the position of the Shine-Dalgarno (SD) sequence and the ATG start codon are all shown on the right. Image taken from *(65)*

4.2 Hfq mRNA Site A, Site B and 5' UTR

To understand how the identified binding sites interact with *E. coli* Hfq three different RNA sequences were used, Site A (5′ – AUUUUUUCAGAAUCGAAAGGUUCA – 3′) which contains site A and the following

stem loop, Site B (5' – GCAUAUAAGGAAAAGAGAGAAUGG – 3') and 5′ UTR which contains all of site A and site B along with the native linker that is between them (Figure 35). Site A and site B were ordered as synthetic oligomers while 5′ UTR was generated using T7 RNA polymerase. To ensure that the proper secondary structures were formed these RNA oligonucleotides were boiled for 5 minutes and then snap cooled on ice. These sequences were then used in the TFQ assay to study their interactions with *E. coli* Hfq. Since these sequences only interact with *E. coli* Hfq the *S. aureus* and *L. monocytogenes* mutants were not utilized in this study.

The site A containing oligoribonucleotide was observed to significantly quench residues Q33W, F39W, F42W, G77W and Y83W. This suggests that site A is capable of binding to both faces with a preference for binding to the proximal face (Figure 37A). The quenching results also indicate that site A quenches two C-terminal tail residues, suggesting a potential interaction between those residues and site A. Y83W was also quenched by $(G\text{-}G\text{-}A)_5$, which bound on the proximal face, potentially suggesting that RNA sequences bound on the proximal face of sufficient length will interact with the tail. The site B oligoribonucleotide was observed to quench F11W, Y25W, Q33W, F39W, F42W, and Y83W (Figure 37B). A preference for distal face binding was observed for site B since the proximal face residues do not quench until the higher 4 μM RNA concentration is reached while the distal face residues quench at the lower 1 μM RNA

concentration. Site B is also observed to quench the C-terminal Y83W mutant indicating that this longer RNA sequence interacts with the C-terminal tail. The behavior of these

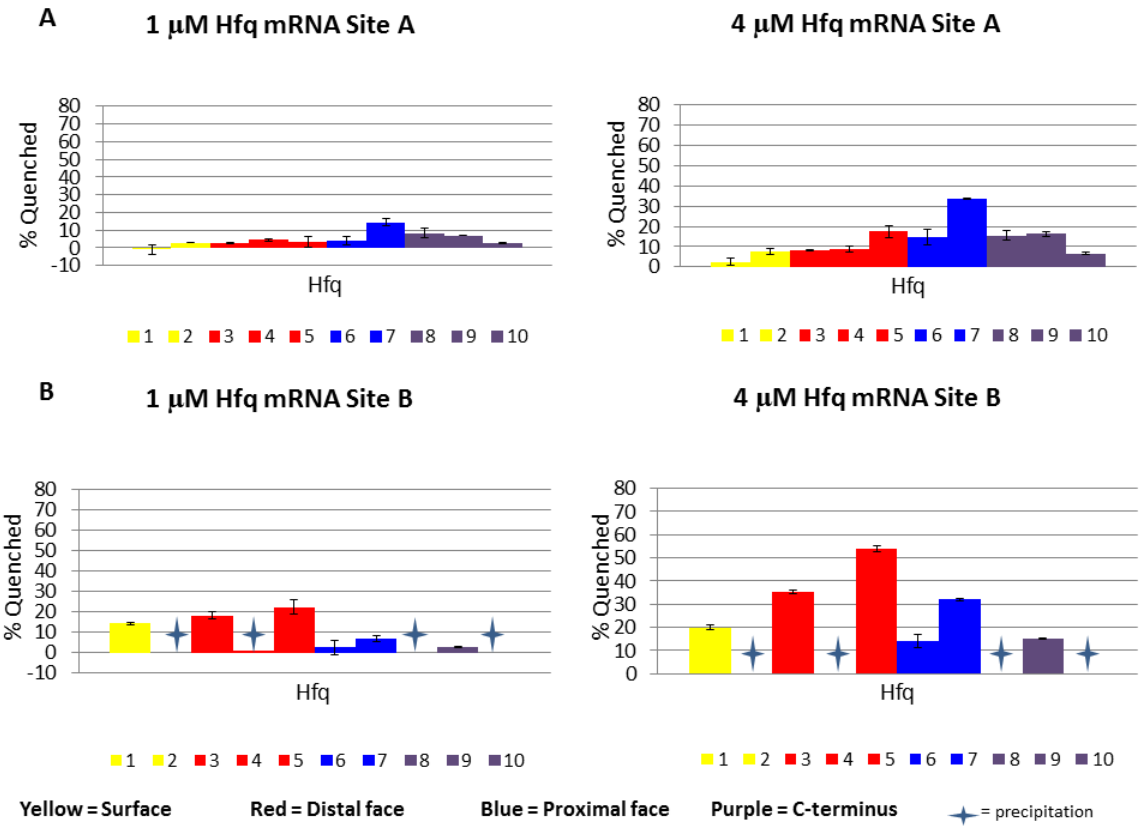


Figure 37: *E. coli* Hfq TFQ using *hfq* site A and site B.

TFQ using the *hfq* mRNA sites known for Hfq binding termed site A and site B. 1 = F11W, 2 = R17W, 3 = Y25W, 4 = K31W, 5 = Q33W, 6 = F39W, 7 = F42W, 8 = G77W, 9 = Y83W, 10 = Q95W. These graphs illustrate that site A can bind to both faces with a preference for binding to the proximal face while site B, which can also bind to both faces, preferentially binds to the distal face.

ligands suggest that it will be possible for both sequences to simultaneously bind to Hfq. To determine if these two sites would compete for a particular face TFQ assays were

done where an initial 1 µM titration of one site was done and then the other site was titrated to a final concentration of 4 µM (Figure 38). These assays were done using

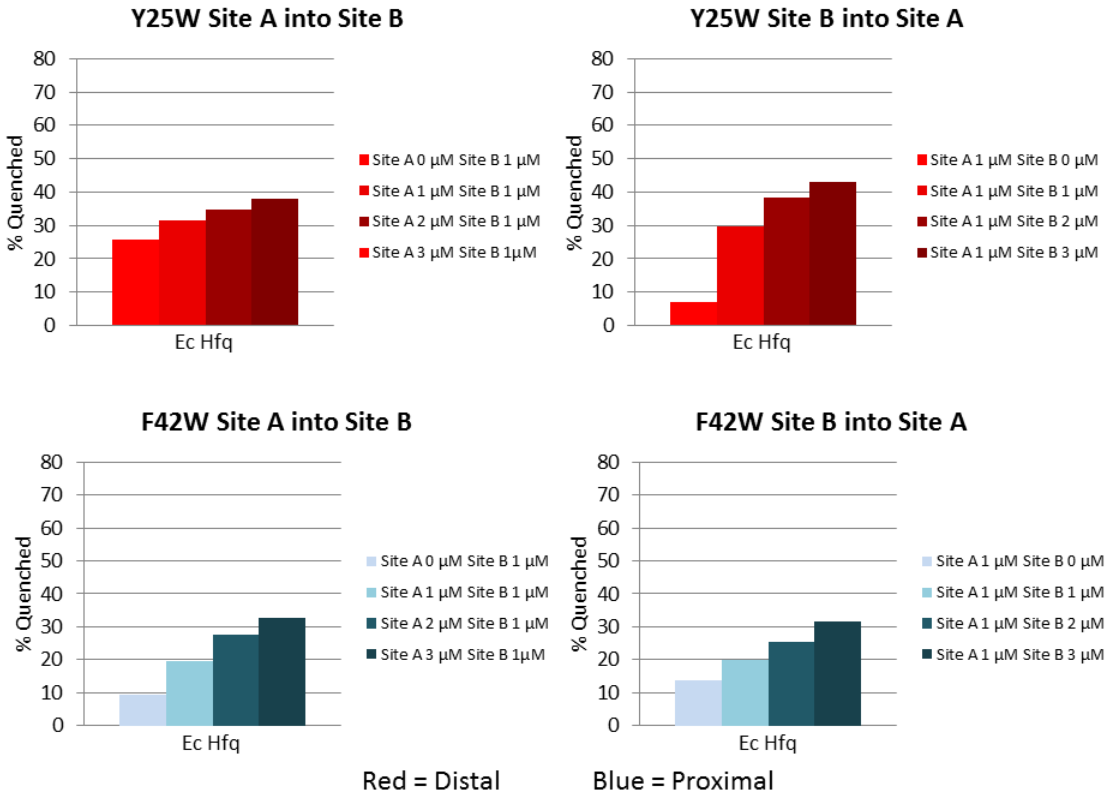


Figure 38: TFQ competition assays using *hfq* mRNA site A and site B

TFQ competition assays using the *hfq* mRNA sites known for Hfq binding termed site A and site B. These graphs illustrate that the two sites do not appear to compete for binding to a particular face.

the full length *E. coli* Hfq Y25W and F42W mutants to analyze binding on each face. For *E. coli* Hfq Y25W it was observed that when Site B was initially titrated followed by Site A there was an initial quench followed by only small increases in quenching as would be expected if Site A were not competing for binding (Figure 38). The reverse titration

where Site A was initially titrated followed by Site B showed a small initial quench followed by more significant quenching also indicating that no competition appears to exist for distal face binding (Figure 38). *E. coli* Hfq F42W was also observed to not experience significant competition for binding (Figure 38). This suggests that the two sequences are not in competition for binding to the faces but rather will bind to opposite faces, most likely binding to the face that each site prefers.

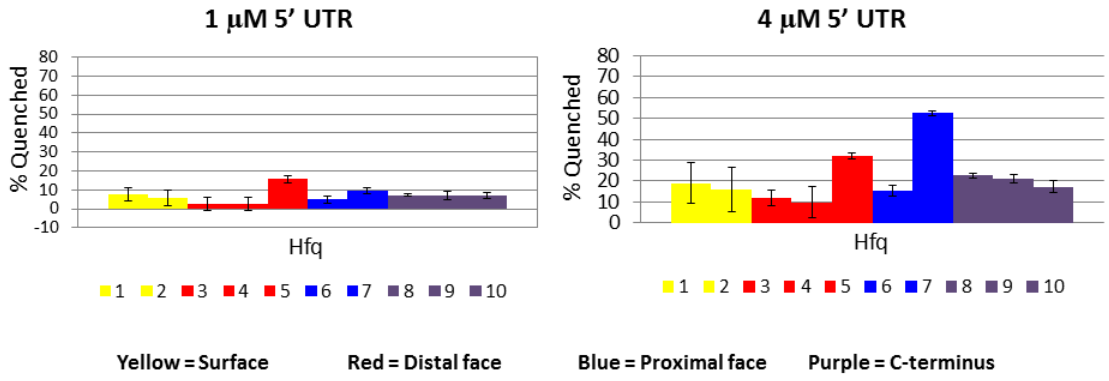


Figure 39: *E. coli* Hfq TFQ using 5′ UTR.

TFQ using the *hfq* mRNA 5′ UTR, which contains both site A and site B. 1 = F11W, 2 = R17W, 3 = Y25W, 4 = K31W, 5 = Q33W, 6 = F39W, 7 = F42W, 8 = G77W, 9 = Y83W, 10 = Q95W. These graphs illustrate that the 5′ UTR binds to both faces with a preference for the proximal face at higher RNA concentrations. These graphs also show that the longer, more structured 5′ UTR is capable of binding the entire C-terminal tail and to the proposed lateral binding site (R17W).

Since there appears to be no competition for binding to the two faces the two sites, site A and site B, were connected using the native linker sequence and that sequence is termed the 5′ UTR. This construct was synthesized biochemically and used

for TFQ (Figure 39). The 5′ UTR sequence quenches all residues to varying extents. Both faces show significant quenching indicating that 5′ UTR binds to both faces with perhaps a preference for the proximal face. The 5′ UTR also quenches all three C-terminal tail residues, G77W, Y83W and Q95W indicating that this significantly longer RNA sequence appears able to interact with the full length of the C-terminal tail. Finally the 5′ UTR quenches R17W which has previously not been observed. R17W resides at the C-terminal end of the N-terminal α-helix on the proximal face. Through our TFQ experiments we have not observed any other interactions with R17W, however, Sauer *et al (52)* and Panja *et al (66)* have proposed a role for residues R16 & R17 in RhyB and *rpoS* binding and have proposed that these residues participate in a third "lateral" binding site. From our observation here it is suggested that longer, more structured RNA sequences may be required for binding to occur at this proposed lateral binding site. This may also suggest that the determined binding mechanisms may not be elucidating the complete binding mechanism for these longer, more structured RNA but rather are illustrating how RNA interacts on a particular face whether it fully or only partially occupies the known binding sites for a given face.

4.3 Proposed Binding Modes of 5' UTR to Hfq

The above results suggest two different potential binding models of the *E. coli hfq* mRNA 5′ UTR to *E. coli* Hfq (Figure 40). The first model (Figure 40A) hypothesizes that

a single 5′ UTR binds to a single Hfq hexamer with site A, which contains the stem loop, binding to the proximal face, site B binding to the distal face and the linker region wrapping around the hexamer thus interacting with the lateral binding site. The second model (Figure 40B) shows two 5′ UTR binding to a single Hfq hexamer where one 5′ UTR is interacting with the proximal face using its site A while the second 5′ UTR is interacting with the distal face via its site B. Both of these models fit the observed data.

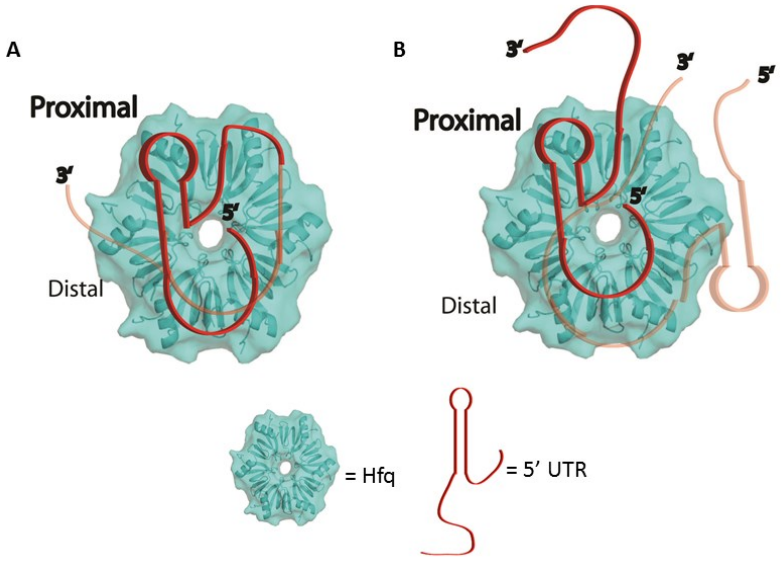


Figure 40: Proposed models for *E. coli hfq* mRNA 5′ UTR binding to *E. coli* Hfq.

Models that satisfy the observed TFQ data for how 5′ UTR interacts with *E. coli* Hfq. A) A single 5′ UTR interacting with the Hfq hexamer B) Two 5′ UTR interacting with the Hfq hexamer.

4.4 Conclusions and Future Directions

TFQ has been used to examine how *hfq* mRNA interacts with *E. coli* Hfq. It is shown that the two known binding sites from the 5′ UTR of *hfq* mRNA do bind to Hfq. Both sequences are capable of binding to both faces however they each show a preference for one face, site A preferring the proximal face while site B prefers the distal face. These two sites do not appear to compete for binding to a particular face, however the competition assay performed here has flaws. To truly know if there is competition between the two sites for Hfq a more thorough competition assay will need to be performed. When the two sites are linked together as found naturally in the 5′ UTR of *hfq* mRNA TFQ is observed to occur on both faces with a potential preference for proximal face binding. The observed TFQ pattern may suggest that site A contributes to binding more than site B, although site B does contribute since total TFQ levels are higher than observed for site A binding alone. Using the 5′ UTR for TFQ also causes all tested C-terminal residues to quench, which has not been previously observed. This indicates that the C-terminal tail residues are involved in RNA binding, however it does not indicate the importance of these residues for binding strength. To determine the relevance of the residues to *hfq* mRNA binding the binding affinity should be determined using alanine point mutants. TFQ with the 5′ UTR also shows quenching of R17W which has not been observed for any other tested RNA sequence. This indicates

that the proposed lateral binding site interacts with RNA. Since TFQ for R17W was observed only for our longest, most structured RNA sequence binding to R17W appears to require longer, structured RNA to interact, which is in agreement with the conclusions from Sauer *et al (52)* and Panja *et al (66)*. Taken together the *hfq* mRNA TFQ suggests two different potential binding modes as shown in Figure 40. The difference between the two models resides around whether or not *hfq* mRNA is binding to Hfq in a 1:1 or 2:1 ratio. There are several techniques that can be used to determine the stoichiometry of binding including gel shift assays, FP, isothermal calorimetry and size exclusion chromatography. Hfq does not required a sRNA to block ribosome binding to *hfq* mRNA thus suggesting the mechanism of action for negative autoregulation follows the known mechanism proposed in Figure 1E where a mRNA binds to *hfq* mRNA, promotes PAP polyadenylating the mRNA and degradation by an exonuclease. The observed binding of Hfq to *hfq* mRNA by TFQ indicates the Shine-Dalgarno sequence will be blocked and that the 3′ end of the RNA will be open for modification and 3′-5′ degradation by an exonuclease.

5. Conclusions and Future Directions

5.1 Conclusions

Hfq is a known RNA chaperone that facilitates sRNA/mRNA pairing. Hfq interacts with many different sRNA/mRNA pairs leading to Hfq regulating many different cellular pathways, including multiple stress responses, quorum sensing, virulence, multidrug resistance and maintenance of membrane integrity. In many bacterial species the deletion of Hfq causes pleiotropic effects including loss of virulence, larger cell size, slower growth, etc *(4-9)*. The mode of RNA interaction with Hfq has been well studied using x-ray crystallography, identifying two binding faces, distal and proximal *(38, 41-46)*. However, these structures all use short RNA sequences that typically are derived from a longer, more complex physiologically relevant RNA. These structures also often use a truncated form of Hfq that removes the highly variable C-terminus. In order to examine how larger RNA sequences interact with Hfq SAXS studies have been performed *(36, 49, 50)*. The SAXS studies suggest that larger RNA strands bind a single face of Hfq *(49, 50)*, however to determine which face those RNA sequences were actually binding other experiments, such as NMR spectroscopy, had to be performed.

Here a quick, straightforward method for analyzing Hfq-RNA interactions has been presented. Using the fact that Hfq natively does not contain Trp it is possible to

use structure guided design to introduce Trp point mutants in Hfq. Since Trp fluoresces it is possible to study Hfq binding to RNA as RNA will quench Trp fluorescence when the two interact. Using known binding sequences, A_{15} and U_6, as control experiments TFQ has been shown to properly identify Hfq-RNA interactions. Specifically it was observed that A_{15} will bind to the distal face of both *E. coli* and *S. aureus* Hfq and U_6 will bind to the proximal face of *E. coli* and *L. monocytogenes* Hfq. To ensure that TFQ occurs only when RNA and Trp directly interact, the Trp residue mimetic NAT was used. No quenching of NAT was observed with either control sequence which demonstrates the necessity of a direct interaction between RNA and Trp to observe quenching.

Further controls were performed to determine what effect the Trp mutants may have on Hfq function and structure. It was observed that a Trp mutation on a binding face lowers binding for RNA sequences that bind to that mutated face, which may lead to false negatives. However it was also observed that these Trp mutants do still bind to their RNA sequences and TFQ is still detected so this lower affinity is not *a priori* detrimental to the interpretation of TFQ data. It is also noted that binding face Trp mutants do not affect binding for RNA sequences that are not expected to bind to that face. Crystal structures were obtained for four of the *E. coli* Trp mutants using the truncated form of *E. coli* Hfq (residues 2-69). Overall these structures demonstrate that mutating residues to Trp does not significantly alter the structure of Hfq. The F11W

structure does vary somewhat from wild type when comparing hexamer:hexamer, which appears to be a result of significant shortening of β strands 2 and 5 caused by the bulky nature of the Trp side chain. β-strands 2 and 5 make up the distal side intersubunit interface and shortening those causes a distortion in hexamer packing leading to a larger RMSD between wild type and F11W. The Y25W structure provides reasoning for the significantly weaker binding to A_{15} (a 184-fold worse binding). In this structure it is seen that the Trp side chain can take two different rotamers, the first rotamer places the indole side chain in the same physical space that adenosine packs into while the second rotamer clashes with the 2'-OH of the RNA substrate. This implies that for the RNA sequence to bind reorganization of the local structure must occur which causes the observed lower binding affinity.

Gram-negative and Gram-positive Hfq homologs have been shown to have different distal face binding motifs. Gram-negative Hfq (*E. coli*) has an $(A\text{-}R\text{-}N)_n$ motif while Gram-positive Hfq (*S. aureus* and *B. subtilis*) have an $(R\text{-}L)_n$ motif. The R-site for both Gram-negative and Gram-positive Hfq has been hypothesized to actually be an A-site due to the fact that all germane crystal structures have adenosine bound in that site, however since it is possible to computationally model a guanosine in the R-site of the *E. coli* crystal structure the sites cannot be defined as adenosine only sites. Binding assays have been performed that indicate Hfq does indeed bind to $(A\text{-}R\text{-}N)_n$ or $(R\text{-}L)_n$

sequences, however, these assays cannot provide insight into where the interaction occurs. Thus, to further investigate the proposed motifs TFQ was applied. TFQ data indicates that the Gram-negative *E. coli* Hfq homologue distal face binding motif is $(A\text{-}A\text{-}N)_n$ and that the Gram-positive *S. aureus* Hfq homologues distal face motif is $(A\text{-}L)_n$. This was determined by using sequences that fulfilled both an $(A\text{-}R\text{-}N)_n$ and $(A\text{-}A\text{-}N)_n$ motif or just an $(A\text{-}R\text{-}N)_n$ motif or sequences that fulfilled either an $(R\text{-}L)_n$ and $(A\text{-}L)_n$ motif or just an $(R\text{-}L)_n$ motif. It was observed that only the sequences that fulfilled both motifs would bind to the distal face while those that fulfilled only the $(A\text{-}R\text{-}N)_n$ or $(R\text{-}L)_n$ motifs will not bind the distal face but may bind the proximal face. The observation that some $(A\text{-}R\text{-}N)_n$ and $(R\text{-}L)_n$ motifs bind to the proximal face is quite interesting and further high resolution structural studies will be pursued to elucidate how these sequences are binding.

Finally TFQ was used to study the interaction of *E. coli* Hfq with *hfq* mRNA. *E. coli* Hfq has been shown to negatively autoregulate its own translation by binding to two sites, site A and site B, within its 5′ UTR. Using TFQ it has been shown that both sites are capable of binding to both faces, however, each site shows a preference for one face. Site A prefers binding to the proximal face while site B prefers the distal face. When the two RNA sequences are joined together via their native linker to create the RNA substrate referred to as 5′ UTR it is observed that the 64 nucleotide sequence binds to

both faces. These data allow two potential models to be proposed that describes how *hfq* RNA binds to Hfq. The first model hypothesizes that a single 5′ UTR binds a single Hfq hexamer such that site A binds to the proximal face, the linker wraps around Hfq and site B binds the distal face. The second model proposes that two 5′ UTRs bind a single Hfq hexamer with one binding to the proximal face via its site A and the other binding the distal face via its site B. These two models can be differentiated by calculating the stoichiometry of binding by ITC or size exclusion chromatography. The 5′ UTR is also observed to bind to the recently proposed lateral binding site. TFQ does not detect binding at this site for any other RNA sequence suggesting that a longer and potentially more structured RNA is required to detect binding here. Finally, binding is observed for the C-terminus out to reside 95 using the 5′ UTR substrate suggesting a longer, more structured RNA sequence will interact with the C-terminus.

5.2 Future Directions

TFQ has been shown to be a useful assay for mapping the RNA binding sites on Hfq. With this technique in hand the next step is to begin analyzing how the vast array of physiologically relevant RNA sequences bind Hfq. Some of these target RNA sequences will include the Spot 42 – *galK* pairing. Spot 42 is a sRNA that downregulates the mRNA *galK*, which encodes the third gene of the galactose operon. Another sRNA/mRNA pair that can be studied is RhyB/*sodB*. RhyB will bind to *sodB* and

downregulates the level of *sodB* in the cell. *SodB* encodes superoxide dismutase, which catalyzes the dismutation of superoxide into oxygen and hydrogen peroxide. There are many different sRNA/mRNA pairs that can be studied and an examination of several different sRNA and mRNA pairs will be performed to understand how these RNA sequences interact with Hfq.

It would be of great interest to expand this method to analyze the proposed "Active cycling" model to analyze how known RNA sequences compete for binding to Hfq. One way to expand TFQ to be able to study the "Active cycling" model is to apply Fluorescence Resonance Energy Transfer (FRET) to TFQ. The goal here would be to examine the current hypothesis that two RNA sequences that bind to the same face on Hfq will actually compete for binding on that face. In this assay the Hfq Trp mutants would be used with two RNA strands. One RNA sequence would have a fluorophore, such as AlexaFluor 350, attached to it while the other RNA sequence will be unlabeled. The Trp mutant would be detected using a spectrophotofluorometer, and then the unlabeled RNA would be titrated in to check for TFQ. Once TFQ has been established the labeled RNA substrate would be titrated into the sample. Excitation would still be targeted at the Trp but now emission will be examined in AlexaFluor 350s range (442 nm). If the RNA strand is capable of competing off the bound substrate FRET will occur, if it is not or if it binds to the opposite face of Hfq, then no FRET will be observed.

Expanding the TFQ assay in this way will provide a technique for directly observing RNA substrates competing for binding to Hfq. The current methods for analyzing RNA cycling on Hfq only indirectly observe which face an RNA sequence is binding to. Currently it has been observed that many RNAs are capable of displacing RNA on either face.

To further validate TFQ, crystal structures of the Trp mutants bound to RNA substrates should be determined, along with solving apo structures for all Trp mutants. Determining these structures will further demonstrate that TFQ properly identifies how an RNA sequence binds to Hfq without altering the known RNA-Hfq interactions.

Studies concerning the distal face binding motif for both Gram-negative and Gram-positive Hfq homologues have raised some interesting questions. The first question that needs follow up relates to how $(G\text{-}G\text{-}A)_5$ binds to *E. coli* Hfq and $(G\text{-}U)_3G$ binds to *L. monocytogenes* Hfq. TFQ indicates that these sequences bind to the proximal face, which was not the expected binding face. These sequences do not have poly-U sequences in them, although the $(G\text{-}U)_3G$ does at least have uridine, which is the nucleotide that typically binds on the proximal face. Therefore to fully understand how these sequences are binding Hfq high-resolution crystallography studies will be pursued.

To date one hypothesis concerning Hfq is that the differential distal face motifs are related to the bacterial species origin of a given homologue. Those that originate from Gram-negative bacteria will display an $(A\text{-}A\text{-}N)_n$ motif while those from Gram-positive will display an $(A\text{-}L)_n$ motif. This is supported by a few crystal structures and the TFQ data presented here. However, the structures and TFQ data each only use a single Gram-negative and two Gram-positive homologs. To truly expand these observed motifs to these groups other Hfq homologs should be used, such as *S. typhimurium* for Gram-negative and *L. monocytogenes* for Gram-positive Hfq. TFQ can easily be applied to other Hfq homologues simply by creating the desired Trp mutants based on sequence alignment and these Trp mutants can then be used to determine if the proposed distal face motifs are truly universal. For Gram-positive Hfq homologues it is useful to study homologs other than *S. aureus* as it is currently controversial whether or not Hfq is important to *S. aureus*. As such Trp mutants have already been created in *L. monocytogenes* using sequence alignment with *S. aureus* to guide the Trp mutant design process. These mutants are currently being taken through TFQ.

Finally, TFQ has demonstrated how the proposed binding sites for the 5′ UTR *hfq* mRNA interacts with Hfq. It has been shown that this RNA is capable of binding to both faces. It is necessary to determine the stoichiometry of binding for the 5′ UTR to Hfq to know if this RNA is capable of wrapping around Hfq as has been proposed. It is

also of interest to design variants of the 5′ UTR that has one of the binding sites abolished. By knocking out one site or the other it can be determined if a particular site is more important for 5′ UTR binding to Hfq.

Appendix

In-cell NMR Spectroscopy Using *E. coli*

Within a cell, there are many different proteins, other biological macromolecules and small molecules that must function properly in order for a cell to survive. The simplest organisms are estimated to use a few hundred small molecules and encode up to 1,000 different proteins *(67, 68)*, with the human genome expected to encode 10-100 fold more *(69)*. The large number and diversity of small molecules and biological macromolecules form a cellular environment that is very crowded and complex *(70)* (Figure 41). Understanding the influence that this packed environment has on protein

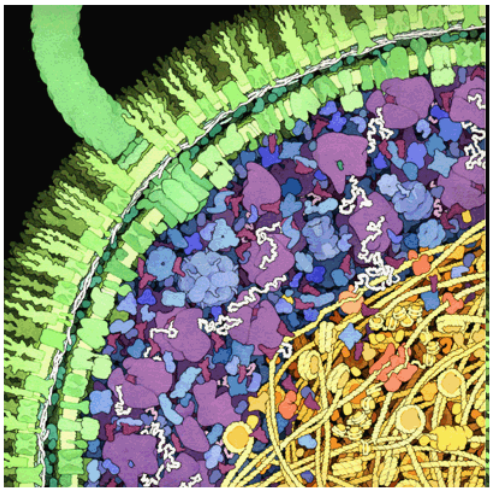


Figure 41: Illustration of the crowded intracellular environment of an *E. coli* cell

This image depicts the current thought about the environment, crowded and complex, that the many known biological macromolecules and small molecules of an *E. coli* cell experience *(70)*.

structure, stability and behavior as well as on protein complexes is one of contemporary biochemical sciences challenges *(67, 71, 72)*. The knowledge that can be gained from complete molecular- and atomic-level observations is invaluable to understanding detailed protein mechanisms of action as well as in the areas of drug development, protein engineering and numerous other research frontiers.

In-cell Nuclear Magnetic Resonance Spectroscopy (NMR) is a technique used to study proteins at atomic-level resolution within living cells *(70)*. The ultimate goals are 3D molecular structure determination and examining binding interactions. Historically, structure determination, both X-ray crystallography and NMR spectroscopy, has been done only on purified proteins. The fact that NMR is a noninvasive solution technique suggests that NMR is uniquely suited to studying full atomic-level macromolecular structures in a representative environment, although not necessarily a natural environment *(72)*. The strength of in-cell NMR spectroscopy lies not only in its ability to determine *de novo* three dimensional (3-D) structures, but also in its ability to observe structural changes of biological macromolecules in a native environment. Structural changes can be directly observed by monitoring changes that occur in a protein two dimensional (2-D) ^{1}H-^{15}N (or ^{1}H-^{13}C) heteronuclear single quantum coherence (HSQC) fingerprint spectrum, since every protein generates a unique ^{1}H-^{15}N (or ^{1}H-^{13}C) HSQC that is dependent on secondary, tertiary and quaternary structure along with the

chemical environment that the protein resides in. Alterations in the structure of the target protein or its chemical environment cause changes in the 2-D fingerprint. This enables studies of molecular interactions within a living cell, such as mapping the binding interface between two proteins *(73)*.

In-cell NMR spectroscopy within *E. coli* has been used to determine the 3-D backbone assignment of GB1*(74)*, determine the *de novo* 3-D structure of gene product TTHA1718 *(75, 76)* as well as to study protein-DNA interactions *(77)*, protein-protein interactions *(73)*, binding events *(78)* and to identify potential new drugs *(79)*. NMR is dependent on nuclei that have nuclear spin such as a proton (^{1}H) whose nuclear spin angular momentum quantum number (*I*) is ½. Other common nuclei used for biochemical NMR spectroscopy include nitrogen (N) and carbon (C), which exist as ^{14}N and ^{12}C isotopes 99.6% and 99.0% of the time respectively, have *I* equal to 1 because they have an even mass number and an even atomic number. An *I* of 1 is NMR inactive as it does not have nuclear spin. Since N and C are the main atoms used for building proteins it is necessary to be able to detect these atoms via NMR, thus the stable isotopes ^{15}N and ^{13}C (occurring 0.4% and 1% of the time, respectively), which have *I* equal to ½, are incorporated into molecules of interest. Both uniform and selective methyl-group labeling strategies have been applied *(80, 81)*. Labeling schemes typically use isotope ^{15}N enrichment for 2-D fingerprint data collection and also incorporate uniform isotope

^{13}C labeling when collecting 3-D data for assignment and structure determination. The ^{1}H-^{13}C HSQC fingerprint can also be useful but it is often complicated by significant background noise arising from incorporation of ^{13}C into cellular metabolites by the cell when the protein of interest is overexpressed within the cell of interest *(81)*. However, it has been demonstrated that utilizing specific ^{13}C methyl-group labeling can be beneficial for observing large proteins that have attenuated ^{1}H-^{15}N HSQC spectra since methyl groups most often have independent rotational motion and give better line shape and intensity *(81-83)*. An alternative specific labeling strategy is to incorporate ^{19}F as an NMR probe, which can enhance sensitivity *(84)*, but typically does not provide sufficient data for structural characterization.

Protein that aggregates or interacts with the cellular membrane, DNA, or large protein complexes may have significantly longer rotational correlation times due to slower molecular motions and/ or may be involved in an intermediate exchange regime. This can lead to NMR signal attenuation due to line broadening, thus making the protein quite difficult to detect and often making the protein "invisible" by in-cell NMR spectroscopy. This was found to be the case for the MetJ repressor protein, which binds nonspecifically to DNA *(77)*.

Recently it has been suggested that the overall charge of a protein will indicate the ability to detect a protein by in-cell NMR spectroscopy where only proteins with an

overall negative charge in the *E. coli* cellular environment, having a pI at or below 4.5, will be detected and proteins that are neutral or positively charged will not be detected *(85)*. This is expected since 40-70% of native *E. coli* proteins are negatively charged along with the cellular environment containing the large negatively charged nucleoid and lipid membrane.

2-D HSQC spectra can be very useful when the *in vitro* assignments for the protein are already known. The fingerprints obtained from the in-cell NMR experiment can often be compared to the *in vitro* spectrum to obtain the assignments unless the structure within the cell is different from the *in vitro* structure. So far, most in-cell NMR HSQC spectra have shown only slight differences when compared to the *in vitro* spectrum, so it is straightforward to transfer the assignments. This also indicates that the molecular crowding and corresponding limited protein-accessible free volume characteristic of the intracellular milieu are not significant determinants of the folded state. With the HSQC spectrum assigned it is straightforward to study how specific changes within the cellular environment affect protein structure. Through examination of 2-D ^{1}H-^{15}N HSQC experiments it is possible to study how drugs bind to the protein of interest within the cellular environment *(78)*, and to identify potential new drugs *(79)*. It is also possible to study protein-protein interactions *(73)* and to study the effect that a post-translational modification on one of the binding partners may have on the binding

interface *(86)*. Even the effect of nonselective protein-DNA interactions on the NMR spectrum has been studied by in-cell NMR spectroscopy, leading to new insights into repressor activity in transcription regulation *(77)*. In-cell NMR spectra have also shown that some, although not all, intrinsically disordered proteins may gain structure within the cellular environment *(87, 88)*. These studies utilizing the 2-D ^{1}H-^{15}N HSQC experiments, illustrate the diversity and strength of in-cell NMR spectroscopy to expand the knowledge of protein function within a living cell.

When 2-D HSQC assignments have not been previously determined in vitro or when significant differences between the two HSQC spectra are observed it becomes necessary to collect additional multidimensional experiments. There are time limitations that can significantly impact the ability to collect multidimensional NMR data as cells do not remain viable, or evenly distributed, within an NMR tube over the time it takes to collect a standard 3-D NMR experiment. These time limitations make it necessary to utilize fast NMR methodologies and ultrasensitive probes that are continuing to be developed and improved.

Cryogenically cooled probes have improved the signal to noise ratio 2-4 fold that of conventional room temperature probes. This improvement is achieved by cooling the RF coils down to 15-30 K, which results in reduced resistance in the coils, lower thermal noise and increased probe quality factor. These probes extend the protein concentration

range that can be detected down to 100 μM or less, or reduce the acquisition time by the sensitivity gain factor squared. Thus cryo-probes aid in detecting protein within a live cell by lowering the required detection limits, however the cryo-probes on their own are not sufficient for collecting multidimensional NMR data using live *E. coli* cells. Thus fast NMR methodologies are also required.

There are several ways to collect fast NMR data *(89-92)*; however, to date only sparse sampling techniques have been used for in-cell NMR spectroscopy. Most fast

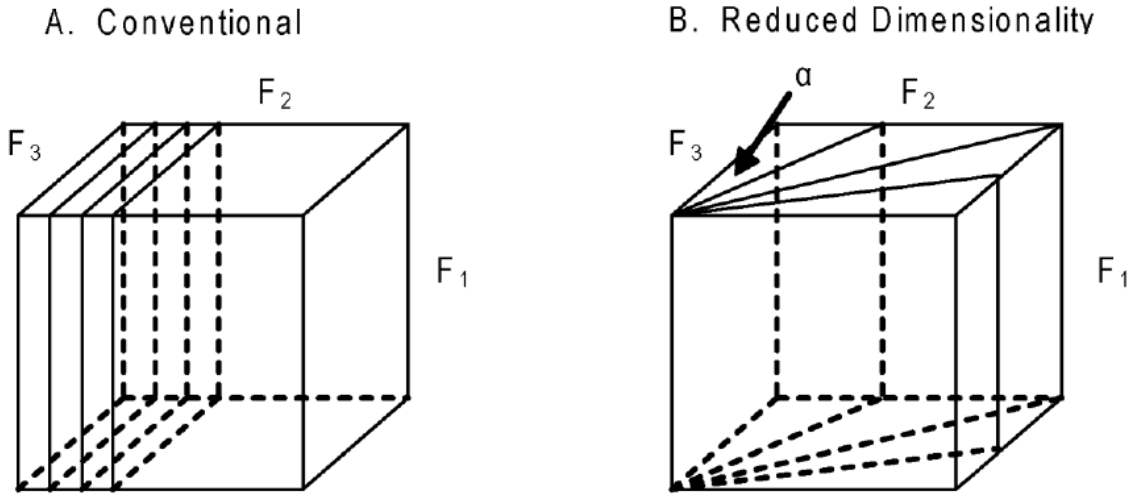


Figure 42: Schematic of the differences between conventional and fast NMR data collection strategies.

F_3 is the directly detected dimension. In a conventional experiment (A), data are collected by holding F_2 to a constant and incrementing in the F_3 dimension. Projection Reconstruction (PR-NMR) (B) has two orthogonal planes, F_1,F_3 and F_1,F_2 with radial projections from the origin formed by the intersection of two orthogonals. The angle α is defined by the rate of incrementation of the indirect evolution times.

NMR methodologies require two aspects to be considered: the sampling pattern used and the processing method applied. The sampling patterns used for fast NMR are designed to reduce the number of points sampled to significantly reduce the data collection time for a specific NMR experiment when compared to the standard Cartesian sampling pattern. Sparse sampling patterns used for multidimensional NMR include a radial sampling pattern *(92, 93)*, concentric ring sampling *(94)* and random sampling *(95)*. Radial sampling is a special case of concentric ring sampling, where the same number of points is taken for each ring (Figure 42). Random sampling involves collecting data points that are distributed randomly. These patterns are different from the standard Cartesian grid sampling pattern as Cartesian sampling distributes the data points equally on a grid while these other sampling patterns either are not on a grid at all or, if positioned to be on a grid, only partially fill it.

The other aspect of performing fast NMR experiments is the ability to properly process the data. There are several processing methods available *(95-102)*. The sampling pattern chosen can lead to artifacts being incorporated in the spectrum upon processing. It is well-known that radial sampling produces artifacts. As a result the field has moved away from radial sampling towards random concentric ring sampling *(103)*. In any of the methods used it is important to appreciate the potential for introducing artifacts in the processed data collected from fast NMR experiments.

Currently two in-cell NMR studies have been performed using these fast NMR techniques. The first study used Projection-Reconstruction NMR (PR-NMR) with radial sampling and the hybrid back-projection/lower-value (HBLV) reconstruction algorithm to walk the backbone of GB1 *(74)*. The backbone assignment of GB1 was accomplished by collecting PR-NMR versions of the 3-D HNCA, HNCO and HA(CA)NH experiments. The second study used random sampling followed by maximum entropy processing *(75)*. The 3-D structure of TTHA1718 was determined using several 3-D heteronuclear NMR experiments and distance restraints obtained from NOE data.

To enhance the understanding of what factors contribute to successful detection of a protein within a living cell by in-cell NMR we pursue the question of how a proteins charge may affect detection by studying how the standard purification tag of 6 His residues affects detection of GB1. We also explore detection of Ubiquitin, which has been reported in the literature but is difficult to reproduce. Currently the largest protein that has been detected by in-cell NMR spectroscopy is calmodulin at 16.8 kDa. To determine size limitations we attempt to observe enhanced green fluorescent protein (EGFP), a 27kDA monomer that has a weak dimerization tendency at concentrations above 5 mg/mL. Finally, to further understand Protein-DNA interactions, Catabolite activator protein will be examined by in-cell NMR spectroscopy with the goal being to study its DNA binding interactions *in vivo*.

Experimental Procedures

Gb1, Gb1-His, Ubiquitin, EGFP and CAP plasmid constructs – The GB1 vector was provided by Dr. Patrick Reardon and selected for using 100 μg/mL Ampicillin. The GB1-his vector was provided by Dr. Pei Zhou's lab and selected for using 25 μg/mL Kanamycin. Ubiquitin was provided by Dr. Patrick Reardon and selected for using 100 μg/mL Ampicillin. Standard mutagenesis methods were employed to create 3A-Ubiquitin. The pGLO vector was provided by Dr. Anne Marie Augustus and selected for using 100 μg/mL Ampicillin. Standard cloning techniques were used to create pET41b-EGFP, which was selected for using 25 μg/mL Kanamycin. The mYPET vector was provided by Dr. Tomoo Ohashi and selected for using 100 μg/mL Ampicillin. The Catabolite activator protein (CAP) was provided by Dr. Charalampos Kalodimos and selected for using 100 μg/mL Ampicillin.

Protein expression for in-cell NMR studies – The desired protein was transformed into *E. coli* BL21 DE3 cells and plated on LB Agar plates containing the appropriate antibiotic as listed above. A 5 mL LB media growth was inoculated using a single colony and grown for 16-20 hours at 37 ^{o}C with shaking at 200 r.p.m. A glycerol stock was saved from the 16-20 hour growth by taking 800 μL of the growth and adding 200 μL of 80% glycerol, flash freezing in a dry ice/ethanol bath and storing at -80 ^{o}C. All subsequent 16-20 hour growths were inoculated from the glycerol stock.

2 50 mL modified M9 minimal media (Table 4) growths were inoculated from a 16-20 hour growth to a starting optical density at 600 nM (OD_{600}) of 0.05. Growths containing the expression vector for either GB1, GB1-His or CAP were then grown at 37 ^{o}C while shaking at 200 r.p.m. to an OD_{600} between 0.4 and 0.6. Cells were then harvested at 1,000 x g for 15 minutes, and then resuspended in fresh modified M9 minimal media containing 1 g/L of $^{15}NH_4Cl$ and 2 g/L ^{12}C-Glucose. Cells were allowed to grow for 10 minutes and then expression of the protein was induced using 1 mM Isopropyl-β-D-1-thiogalactopyranoside (IPTG) and grown at 37 ^{o}C with shaking at 200 r.p.m. for 4-6 hours. After 4 hours of expression one growth was harvested at 1,000 x g for 15 minutes at 4 ^{0}C. The cell pellet volume was estimated using known volumes of water and then the pellet was resuspended to a 20 % cell slurry in modified M9 minimal media containing 10% D_2O. This sample was used to tune and shim the NMR instrument. After 6 hours of expression the second growth was harvested at 1,000 x g for 15 minutes at 4 ^{o}C. This sample was resuspended in the same manner as the first sample. This second sample was then used to collect the desired NMR experiment. If protein signal was detected then the sample was removed from the NMR tube, the cells spun out at 1,000 x g for 10 minutes and the supernatant placed into the NMR tube. The exact same NMR experiment was collected on the supernatant to ensure that the observed protein signal resides within the cells and not outside the cells. If protein

Table 5: Modified M9 Minimal Media Composition

Component	Concentration
Sodium Phosphate (Dibasic)	80 mM
Potassium Phosphate (Monobasic)	40 mM
NaCl	8.6 mM
$MgSO_4$	2 mM
$FeCl_3$	1 μM
$ZnSO_4$	25 μM
$CaCl_2$	100 μM
H_3BO_3	2 μM
$CuSO_4$	2 μM
$CoCl_2$	2 μM
$MnCl_2$	10 μM
$NiSO_4$	2 μM
$(NH_4)_6Mo_7O_{24}$	2 μM
$(NH_4)Cl$	0.1% (w/v)
Glucose	0.2% (w/v)

signal was not detected then the cells were lysed by freeze-thaw and the cleared lysate was examined for protein signal to see if protein expression levels were sufficient for detection.

Expression for EGFP, mYPET and ubiquitin was done by growing 2 50 mL modified M9 minimal media samples at 37 ^oC with shaking at 200 r.p.m. until OD_{600} reached between 0.4 and 0.6. Expression was then induced using 1 mM IPTG and cells

were grown at 15 ^{0}C for 16-20 hours. The two samples were prepared and used in the same manner as those for GB1, GB1-his and CAP.

Freeze-thaw cell lysis – If no protein signal was detected after the desired NMR experiment is collected then the cell slurry is removed the NMR tube, the cells harvested and resuspended in an equal volume of lysis buffer (50 mM Phosphate buffer pH 7.0, 300 mM NaCl, 2 mM PMSF, 2 µg/mL Deoxyribonuclase I (DNase I), 1.5 mg/mL lysozyme). Cells were then frozen in a dry ice/ethanol bath for 5 minutes, removed and allowed to thaw at room temperature. Cells were frozen and thawed two more times. The insoluble debris was spun out at 16,000 x g for 10 minutes at 25 ^{o}C. The supernatant was removed, 10% (v/v) D_2O added and the desired NMR experiment collected.

NMR data collection – NMR spectra were collected on either a Varian 600 MHz or 800 MHz Innova spectrometer equipped with a H,C,N cryogenically cooled triple resonance probe. All spectra were processed with nmrPipe and visualized with nmrView *(104, 105)*.

CAP purification – *E. coli* cells expressing WT-CAP were lysed using Bugbuster (EMD Millipore, Billerica, MA) containing 25 µg/mL DNase following the standard protocol. The lysate was then spun at 15,000 x g for 30 minutes at 4^{o}C. The clarified lysate was then loaded onto a prepared nickel column. The nickel column was then washed using buffer A (50 mM potassium phosphate^{-} pH 8.0, 300 mM KCl and 20 mM Imidazole) for 5

CV's. Next a gradient elution was setup to go from 100% Buffer A to 100% Buffer B (50 mM potassium phosphate pH 8.0, 300 mM KCl, 250 mM Imidazole) in 30 minutes at a flow rate of 1 min/mL. During elution 2 mL fractions were collected and monitored for absorbance at 280 nm. The fractions containing WT-CAP were then concentrated and buffer exchanged into 50 mM potassium phosphate pH 6.0, 500 mM KCl, 1mM DTT. WT-CAP was concentrated to 1 mM.

CAP binding to sonicated salmon sperm DNA with and without DNase – CAP was used at a concentration of 500 μM for all NMR experiments. An initial ^{1}H-^{15}N HSQC was collected and then sonicated salmon sperm DNA was titrated incrementally to be at 50 μg, 100 μg and 250 μg total DNA. At each point a new ^{1}H-^{15}N HSQC was collected. To test the role of cAMP in DNA binding a 500 μM CAP sample containing 250 μg of DNA was used. cAMP was titrated in going upto 1 mM CAP (3.37 μM, 4.35 μM, 9.4 μM, 50 μM, 250 μM, 500 μM and 1 mM). A ^{1}H-^{15}N HSQC was collected at each point. Finally to perform the clarified lysate experiment BL21 DE3 *E. coli* cells were grown in 5 mL LB media overnight. The cells were then harvested at 4,000 x g for 10 minutes. These cells were then lysed using Bugbuster and the lysate was clarified at 15,000 x g for 30 minutes at 4 ^{o}C. WT-CAP was then added to 700 μL clarified lysate to a concentration of 500 μM. To digest the genomic DNA 25 μg/mL DNase was added to the sample and the sample was incubated for 10 minutes at 25 ^{o}C before collecting a ^{1}H-^{15}N HSQC.

In-cell NMR Spectroscopy of GB1 within E. coli

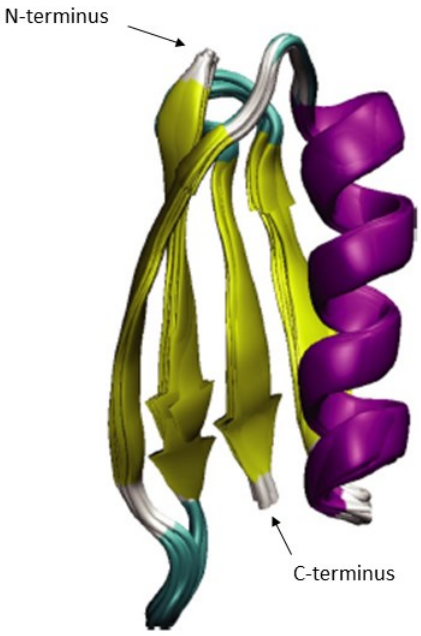


Figure 43: Structure of GB-1 solved by NMR spectroscopy

The *in vitro* structure of GB-1. It forms a 4 stranded β-sheet where the inner two strands are parallel while the outer two are antiparallel with a α-helix on one side as shown. B-strands are colored yellow while the α-helix is colored purple.

The B1 domain of protein G (GB1) is a stable domain that binds the F_c of immunoglobulin G (IgG) *(106)*. This domain has 56 amino acids and is about 7 kDa. It forms a 4 stranded β-sheet with a single α-helix (Figure 43). The two inner β-strands are parallel while the outer two β-strands are antiparallel. Its stability has led to this protein becoming an NMR standard. Thus this protein has been used for in-cell NMR

spectroscopy in *E. coli (74)*, *X. laevis* oocytes *(107)*and HeLa *(108)* cells. GB1 was used in *E. coli* for the first successful 3-D backbone assignment by NMR *(74)*.

Recently it has been proposed that only proteins that have an overall negative charge within the cell, by having a pI that is significantly lower than the native pH of the cytosol of the cell, which is around 7, will be visible by in-cell NMR spectroscopy within living *E. coli* cells *(85)*. This was reasoned to be the case due to the fact that 40-70% of native *E. coli* cells are also negatively charged along with the presence of the negatively charged nucleoid and cellular membrane. Thus any negatively charged protein would be repulsed and thus tumbling relatively freely in solution while those proteins that are either uncharged or positively charged would have transient interactions with the negatively charged components of the cell thus causing them to have large apparent molecular weights and become invisible by NMR. GB1 was used in this study as the example of the protein that has an overall negative charge, since its pI is 4.37, and is thus successfully detected within a live *E. coli* cell.

To investigate this intriguing claim that only negatively charged proteins will be detected by in-cell NMR spectroscopy we placed a c-terminal 6-His tag onto GB1. The 6-His tag raises the pI of GB1 to 5.54. Based on the ranges reported *(85)* this classifies GB1 as being a neutral charged protein and it should thus no longer be detected within a living *E. coli* cell by NMR. We grew cells expressing either GB1 or GB1-His and

collected ^{1}H-^{15}N HSQC's (Figure 44). The untagged GB1 was easily detected by in-cell NMR spectroscopy while the GB1-His was not detected (Figure 44). To verify that the inability to detect GB1-His was not due to poor protein expression the cells were lysed by sonication, the insoluble debris removed and a ^{1}H-^{15}N HSQC was collected. There was sufficient expression for detection by NMR (Figure 44). These results suggest that indeed the charge of the protein will affect the ability to detect a protein within a living *E. coli* cell. These results also suggest that rather than the overall charge of the protein affecting in-cell NMR detection a localized charge, as the 6-His tag on GB1 really is a localized positive charge, can affect detection. This observation would suggest that if a protein were designed in such a way that the surface it presented to the cell were negatively charged detection may become possible. Further studies would include adding other tags to the terminus of GB1 that span the range of charges, such as adding a 6-Lys tag, a 6-Ala tag or a 6-Asp tag to have a positive, neutral and negatively charged tag. The detection or lack of detection would aid in understanding how charge affects in-cell NMR detection. Finally, proteins that are known to be invisible, such as ubiquitin or FKBP, should be designed to present themselves as negatively charged and in-cell NMR experiments collected.

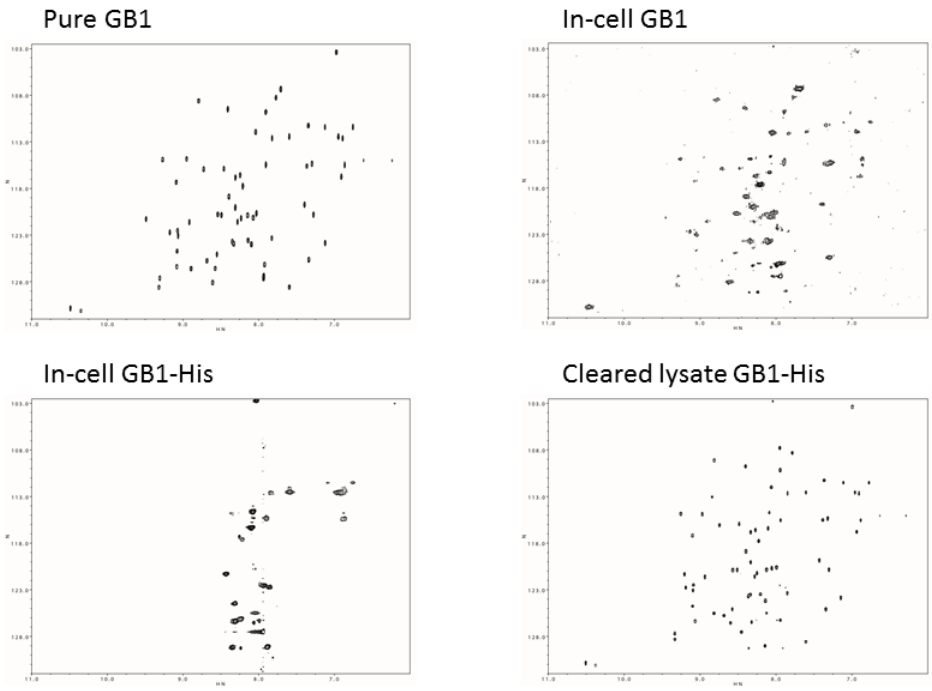


Figure 44: In-cell NMR Spectroscopy using GB1-His

Various ^{1}H-^{15}N HSQC spectra of GB1. The top left spectrum illustrates the fingerprint of GB1 *in vitro* as a reference spectrum for the in-cell NMR spectra. The top right spectrum was collected on *E. coli* cells expressing GB1 while the bottom left spectrum was collected on *E. coli* cells expressing GB1-His. The bottom right spectrum was collected using the cleared lysate from the sample that was used to collect the bottom left spectrum illustrating that expression level of GB1-His was sufficient for detection by NMR.

In-cell NMR Spectroscopy of Ubiquitin within E. coli

Ubiquitin is a 76 residue, 8 kDa protein that is ubiquitous in almost all eukaryotic cells. Ubiquitin is best known for its role as the modification to other proteins that indicates that the tagged protein should be taken to the proteasome for degradation.

This degradation signal requires that polyubiquitin chains, consisting of at least 4 ubiquitins, are formed through lysine 48 linkage of ubiquitin. Ubiquitination can also signal many other pathways via either lysine 63-linked polyubiquitination or

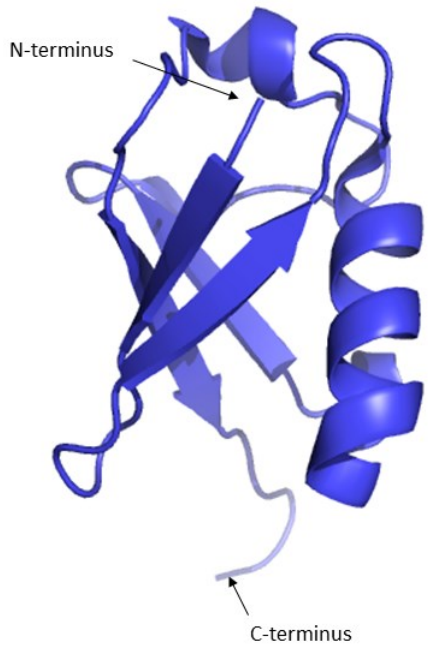


Figure 45: Crystal Structure of Ubiquitin

The *in vitro* structure of Ubiquitin. It consists of a mixed β-sheet, one α-helix and a short piece of 3(10) α-helix. Image was generated in Pymol using PDB ID: 1UBQ *(109)*.

monoubiquitination. Some of the cellular processes that ubiquitin is involved with include: endocytosis, DNA repair, vesicular trafficking and signaling pathways. The overall structure of ubiquitin consists of a mixed β-sheet, one α-helix and a short piece of 3(10) α-helix (Figure 45) *(109)*. While ubiquitin is most often thought of as interacting

with proteins via covalent linkages, ubiquitin can also associate with proteins via non-covalent interactions such as hydrophobic interactions. These non-covalent interactions are used by two of the three classes of enzymes that are responsible for ubiquitination, the E1 and E2 families. The E3 family is covalently linked to ubiquitin and is responsible for transferring this linkage to the target protein. Non-covalent interactions occur between ubiquitin and other proteins via the hydrophobic patch that ubiquitin presents on the face of the mixed β-sheet that is not blocked by the α-helix (Figure 46). Ubiquitin serves many functions within the cell and along with its small size has been identified as an interesting protein to study using in-cell NMR techniques. To date ubiquitin has been reported to be detected by NMR spectroscopy within *E. coli* cells *(73)*, *Pichia pastoris* (*P. pastoris*) *(110)* and HeLa cells *(108)*. To observe ubiquitin within HeLa cells the residues that comprise the hydrophobic patch were mutated to Ala. These residues had to be mutated in order for ubiquitin signal to be detected in HeLa cells by NMR spectroscopy *(108)*. The observation that ubiquitin can be detected within *E. coli* cells is currently an observation that only Burz et al *(73)* can make, other labs that have attempted to detect ubiquitin by in-cell NMR spectroscopy within *E. coli* have not been successful. Gierasch et al *(85)* suggest that ubiquitin, which has a pI of 6.56, is neutrally charged within the cellular environment and thus will not be detected by in-cell NMR spectroscopy, which is what they observed *(85)*.

Ubiquitin has many interesting biological roles and being able to detect ubiquitin within live cells by NMR spectroscopy would enhance the understanding of how

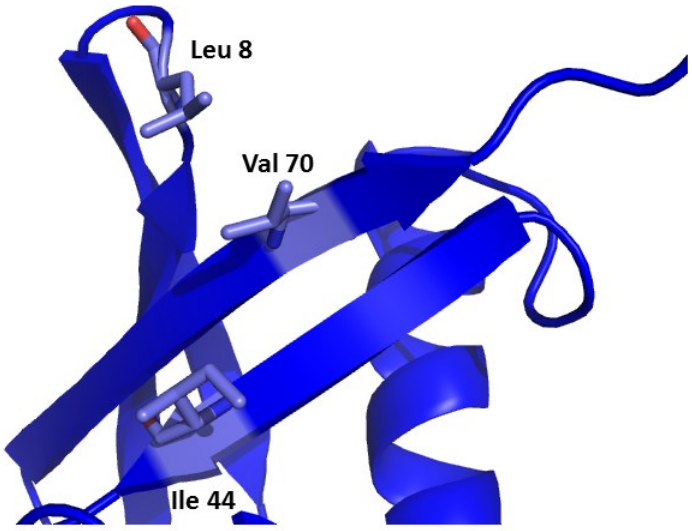


Figure 46: Zoom in of the hydrophobic patch of ubiquitin

Focus on the hydrophobic patch responsible for most non-covalent interactions that occur between Ubiquitin and its many binding partners. Image generated in Pymol using PDB ID: 1UBQ *(109)*.

ubiquitin functions along with displaying the versatility that in-cell NMR spectroscopy has. We therefore attempted to determine what is necessary to successfully detect ubiquitin within living *E. coli* cells. We were incapable of detecting wild type ubiquitin within *E. coli* cells. By following the exact protocol reported by Shekhtman et al *(73)*, which includes freezing the cells after expressing protein but before collecting the

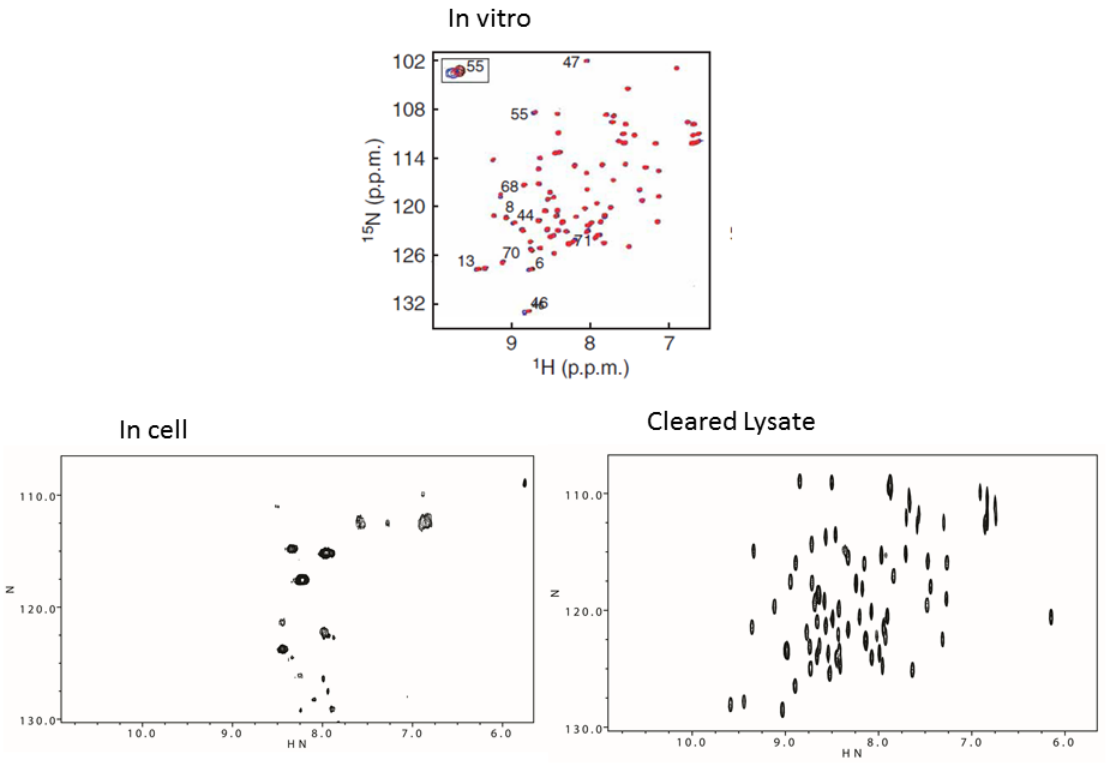


Figure 47: ^{1}H-^{15}N HSQC Spectra of 3A Ubiquitin

^{1}H-^{15}N HSQC spectra of ubiquitin. The top spectrum shows the ubiquitin spectrum from an *in vitro* sample while the bottom left shows a spectrum of ubiquitin collected using live *E. coli* cells that are expressing ubiquitin. The bottom right panel shows the spectrum obtained after lysing the cells used to collect the bottom left sample.

desired NMR experiment, a spectrum of ubiquitin can be obtained. Freezing cells is a well-known cell lysis technique and in our hands we detected ubiquitin in the supernatant however, Shekhtman et al report that their supernatant is clear.

To detect ubiquitin within living HeLa cells by NMR Inomata et al *(108)* had to mutate the three residues that make the hydrophobic patch on ubiquitin to Ala. We created the same triple mutant of ubiquitin, expressed the protein and collected a ^{1}H-^{15}N HSQC. We did not detect any signal from ubiquitin (Figure 47). We then lysed the cells from that sample and collected a ^{1}H-^{15}N HSQC on the cleared lysate. The expected spectrum of ubiquitin was observed indicating that sufficient expression of ubiquitin was achieved, thus suggesting that an interaction within the cell is causing signal attenuation of ubiquitin.

Finally, it has been shown that ubiquitin will aggregate in the presence of copper *(111)*. The concentration of copper in the modified M9 minimal media is sufficiently high that it could potentially be causing ubiquitin to aggregate within the cell thus causing a large apparent molecular weight and signal attenuation. 3A ubiquitin was therefore grown and expressed in modified M9 minimal media that had no copper and used to collect a ^{1}H-^{15}N HSQC. No ubiquitin signal was detected however ubiquitin was detected in the cleared lysate thus indicating that sufficient expression was achieved (Figure 48). These results suggest that the ability to detect ubiquitin within live *E. coli* cells is not well understood and further studies of the exact parameters for reproducing Shekhtman et al's results need to be conducted. The final difference between the experiments resides with the promoter system that has been used.

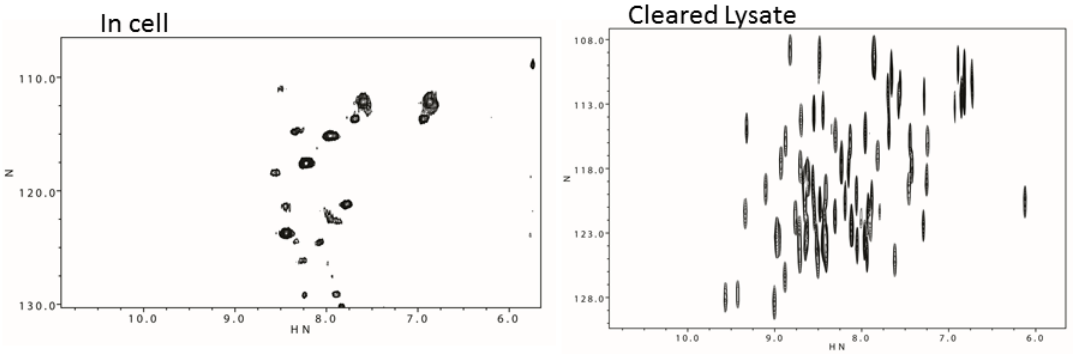


Figure 48: ^{1}H-^{15}N HSQC spectra of 3A ubiquitin grown in M9 minimal media lacking copper.

^{1}H-^{15}N HSQC spectra of ubiquitin. The left shows a spectrum of ubiquitin collected using live *E. coli* cells that are expressing ubiquitin and have been grown in minimal media lacking copper. The right panel shows the spectrum obtained after lysing the cells used to collect the bottom left sample.

For our studies we induced expression using the T7 promoter while Shekhtman et al used the arabinose promoter. The arabinose promoter tends to be a weaker expression system thus it is possible that the T7 promoter system is producing too much protein, thus causing non-specific interactions between ubiquitin monomers leading to signal attenuation. Finally, this also suggests that careful controls must be done and that all in-cell NMR experiments should be include a cell viability assay to ensure that the majority of cells are viable. It is highly recommended that cells intended to be used for performing in-cell NMR experiments not be frozen prior to collecting the in-cell NMR experiment as this can lead to cell lysis and protein being present in the supernatant.

In-cell NMR Spectroscopy of EGFP within E. coli

Green Fluorescence Protein (GFP) is a 238 amino acid, 27 kDa protein that was originally identified from the jellyfish *Aequorea aequorea (112)*. This protein will fluoresce green when it is excited by blue light. In contrast to most fluorescent and luminescent

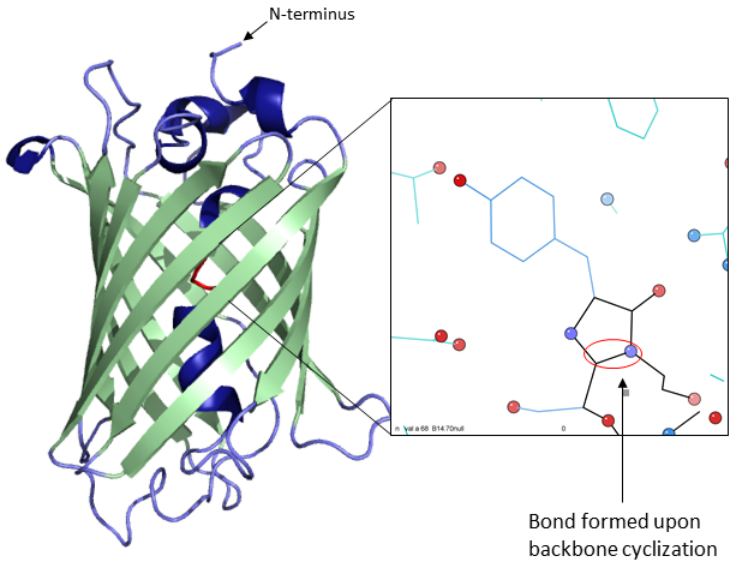


Figure 49: GFP X-ray structure

The structure of GFP, a β-barrel consisting of 11 β-strands and six α-helices. The chromophore, 4-(p-hydroxybenzylidene)imidazolidin-5-one (HBI), resides between the 2nd and 3rd α-helices. The zoomed in window illustrates where the backbone cyclizes to form HBI. Image of GFP created in Pymol using PDB ID: 1GFL *(113)*.

proteins GFPs fluorescence capability originates from the protein structure rather than an acquired cofactor. GFP forms a β-can structure with six α-helices *(113)*. Residues Ser65 and Gly67, which reside between the second and third α-helices, will cyclize along

the backbone and Tyr66 gets oxidized to form the chromophore, 4-(p-hydroxybenzylidene)imidazolidin-5-one (HBI) (Figure 49) *(114)*. Enhanced GFP (EGFP) is a variant of GFP that has three point mutations, F100S, M154T and V164A *(115)*. EGFP is commonly used in cellular biology as a tag to localize other proteins via fluorescence microscopy techniques. EGFP has been shown through many cellular biology studies to not interact with other molecules within a cell, thus allowing the conclusion that where fluorescence from EGFP is observed will be due to the localization of the protein that EGFP has been fused to *(114)*. Since EGFP does not interact with the components of the cell or with other proteins it should be favorable for 3-D structural studies of a higher molecular weight protein by in-cell NMR spectroscopy.

EGFP was overexpressed in *E. coli* BL21 DE3 cells using the pGLO vector (Bio-Rad, Hecules, CA) and the arabinose promoter. EGFP was expressed overnight at 15 ^{o}C to ensure proper folding in ^{15}N modified M9 minimal media. A ^{1}H-^{15}N HSQC was collected. No signal for EGFP was detected (Figure 50) so the cells were lysed and a ^{1}H-^{15}N HSQC collected however there was still no signal detected from EGFP. This suggests that either the concentration of EGFP is not sufficient, EGFP is dimerizing with itself (EGFP has a weak dimerization tendency at concentrations above 5 mg/mL) and the 54 kDA dimer weight is leading to signal attenuation or EGFP is interacting with other cellular components that are causing a large apparent MW and thus leading to

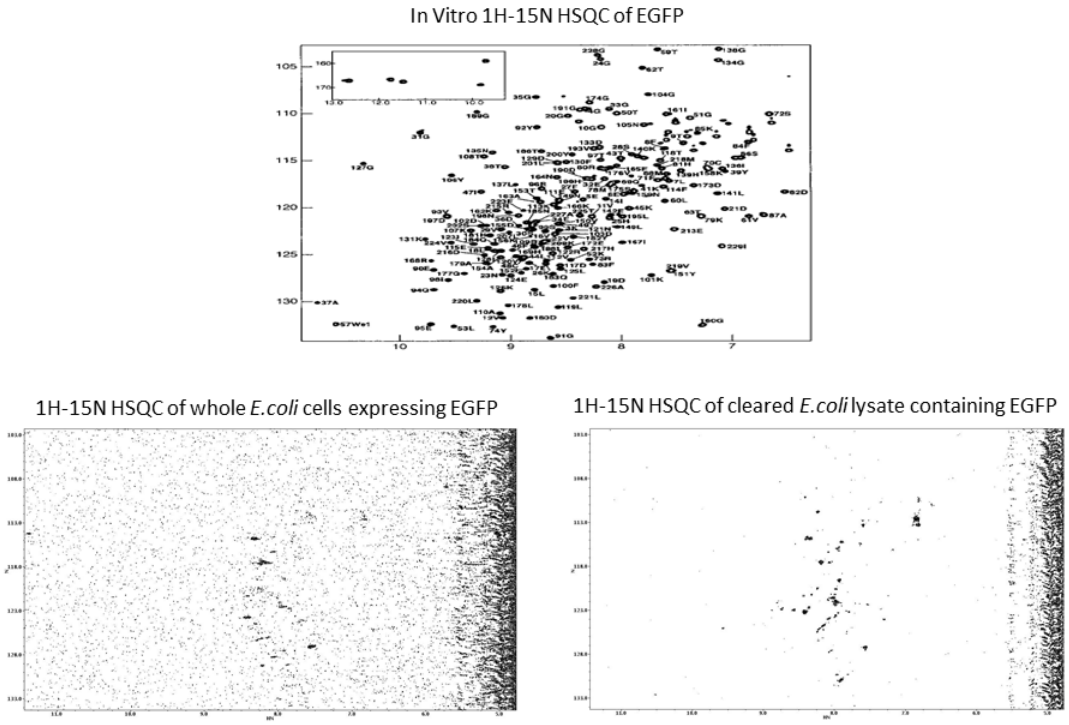


Figure 50: ^{1}H-^{15}N HSQC spectra of EGFP

^{1}H-^{15}N HSQC spectra of EGFP. The top spectra shows the *in vitro* ^{1}H-^{15}N HSQC spectra for EGFP *(116)*.The bottom left shows a spectrum of EGFP collected using live *E. coli* cells that are expressing EGFP. The bottom right shows the spectrum obtained after lysing the cells used to collect the bottom left sample.

signal attenuation. The tendency to dimerize can be abolished with a single point mutant, A206K. Thus a variant of EGFP, mYPet, was kindly provided by Dr. Harold Erickson's lab. This variant is induced using the T7 promoter system, which is a stronger promoter than the arabinose promoter, and should express greater amounts of mYPet compared to EGFP. The same expression protocol and NMR experiment was

performed and once again no signal was detected (Figure 51). Upon cell lysis no mYPet signal was detected (Figure 51). This suggests that either the protein concentration is

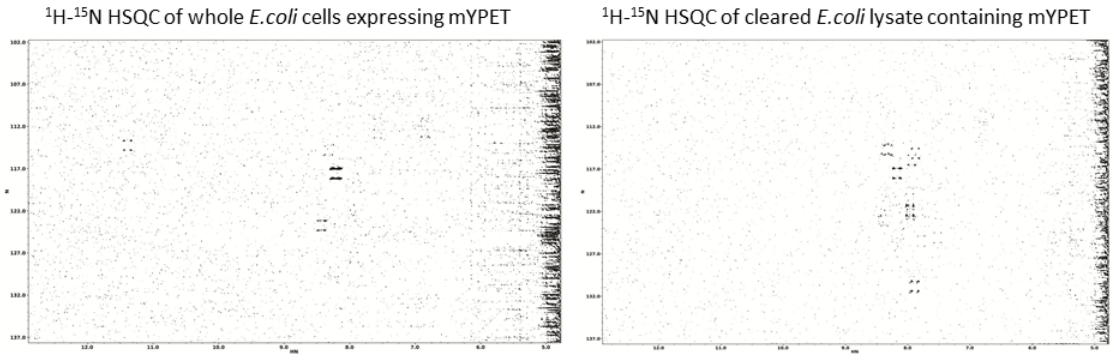


Figure 51: ^{1}H-^{15}N HSQC spectra of mYPet

^{1}H-^{15}N HSQC spectra of mYPet. The left panel shows a spectrum of mYPET collected using live *E. coli* cells that are expressing mYPET. The right panel shows the spectrum obtained after lysing the cells used to collect the left sample.

still too low for detection by NMR or that GFP is interacting with some cellular component leading to signal attenuation.

To ensure that GFP was indeed expressing and not localizing within the cell fluorescent images of *E. coli* cells expressing EGFP were taken using a fluorescent microscope. These images illustrated that EGFP was indeed expressing, folding correctly and not localizing within the cell (Figure 52). Considering the standard use of attaching GFP to proteins *in vivo* and using the fluorescence of GFP to localize where the other protein is found within a cell and the fairly low expression levels detected on an SDS-PAGE gel the most likely explanation is that the expression level of EGFP is too low

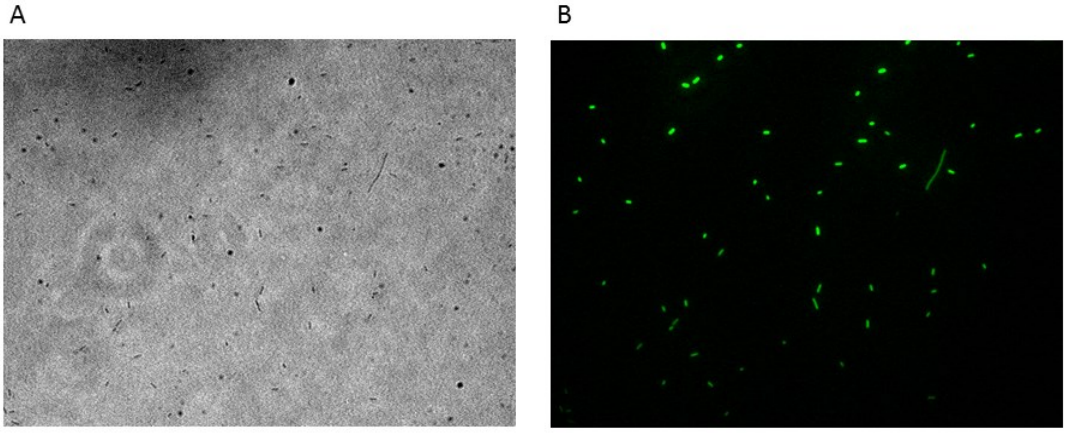


Figure 52: Microscopic images of *E. coli* expressing EGFP

Microscopic images taken with a fluorescent microscope. A) A differential interference contrast (DIC) image of *E. coli* cells B) Fluorescent image of the same cells as in A.

for in-cell NMR detection. Therefore, to observe GFP within live *E. coli* by in-cell NMR spectroscopy the expression levels of GFP will need to increase. Varying the amount of IPTG used for the mYPet expression and increasing the length of time the cells are grown before harvesting are two methods of optimizing protein expression that have not yet been pursued.

In-cell NMR Spectroscopy of CAP within E. coli

Catabolite activator protein (CAP) is a 209 amino acid, 22.5 kDa protein native to *E. coli* that functions as a symmetric homodimer. CAP consists of two domains, the DNA binding domain and the cAMP binding domain (Figure 53). CAP is a transcriptional

activator that interacts with over 100 different promoter regions by binding to its consensus DNA recognition sequence, 5′ - ATGTGATCTAGATCACATTT – 3′. CAP

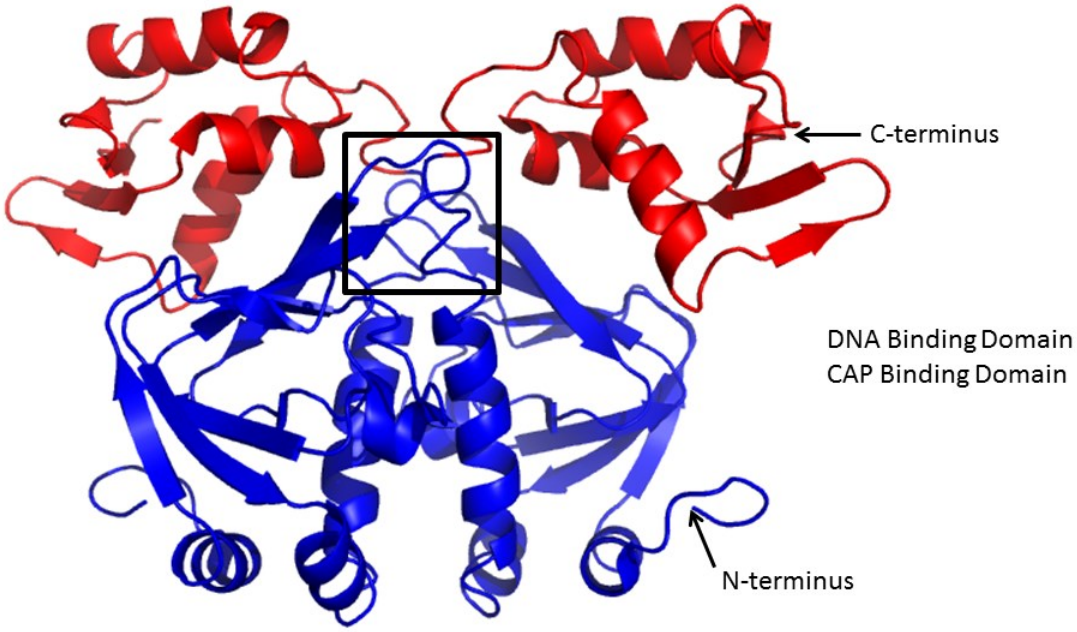


Figure 53: Apo Catabolite Activator Protein (CAP) NMR structure

The structure of CAP revels that there are two domains, the cAMP-binding domain (blue) and the DNA binding domain (red). When cAMP binds the coil region, shown within the black box, will clamp down on cAMP which causes an α-helix leading to a 60° rotation of the DNA binding domain, thus allowing α-helix F to align with the major grooves *(117)*. This image was made in Pymol using PDB ID: 2WC2 *(117)*.

utilizes 3′- 5′ adenosine monophosphate (cAMP) for its mechanism of action. When CAP is in its apo form it does not bind to DNA. However, upon binding of cAMP to CAP the c-terminal domain rotates about 60° with respect to the N-terminal domain thus allowing CAP to bind to its specific DNA recognition sites near promoter regions

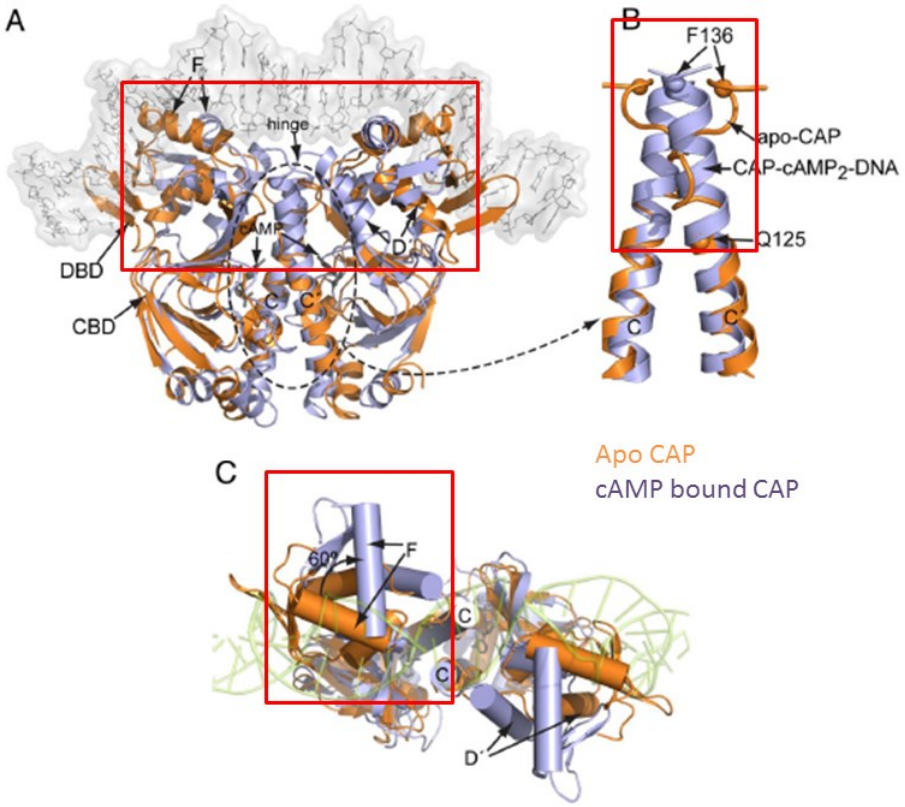


Figure 54: Comparison of Apo CAP and cAMP-CAP Structures

Overlay of the Apo (Orange) and cAMP (Light purple) bound structures of CAP. A) Overview of entire complex illustrating that in the Apo form α-helix C is unfolded at its c-terminus compared to the bound structure. This unfolding leads to a 60° rotation of α-helix F causing α-helix F to not be properly aligned for binding to DNA. B) Zoom on α-helix C. C) Zoom on α-helix F. The red boxes are illustrating where the differences between the two structures reside. *(117, 118)*

(Figure 54). Once CAP has bound to DNA it will enhance the ability of RNA polymerase to bind with DNA and enhance transcription. If this model is indeed the case then it should be possible to observe CAP in its apo form within live *E. coli* cells using in-cell NMR spectroscopy and then supply cAMP externally to the *E. coli* and

monitor changes in the ^{1}H-^{15}N HSQC spectra as the cAMP is taken up by the *E. coli* cells and CAP binds to DNA. A similar in-cell NMR study was attempted using the methionine transcriptional repressor protein, MetJ. This protein represses the met regulon in the presence of S-adenosylmethionine (SAM). In this study it was not possible to detect MetJ within a living *E. coli* cell. It was determined that the cause of signal attenuation is due to a constant non-specific interaction of MetJ with DNA which leads to an increase of the apparent molecular weight for MetJ *(77)*.

The first experiment performed involved determining whether or not pure CAP would interact with non-specific DNA. For this experiment pure ^{15}N-labeled CAP was generated and used at a concentration of 721 μM. A ^{1}H-^{15}N HSQC was collected. Next 50 μg of sonicated salmon sperm DNA was added to the sample and another ^{1}H-^{15}N HSQC spectra was collected. This process was repeated iteratively until 250 μg of sonicated salmon sperm had been added (Figure 55). This experiment demonstrated that when CAP was in its apo form it would not bind in a significant manner to nonspecific DNA thus suggesting that an in-cell NMR experiment may be feasible.

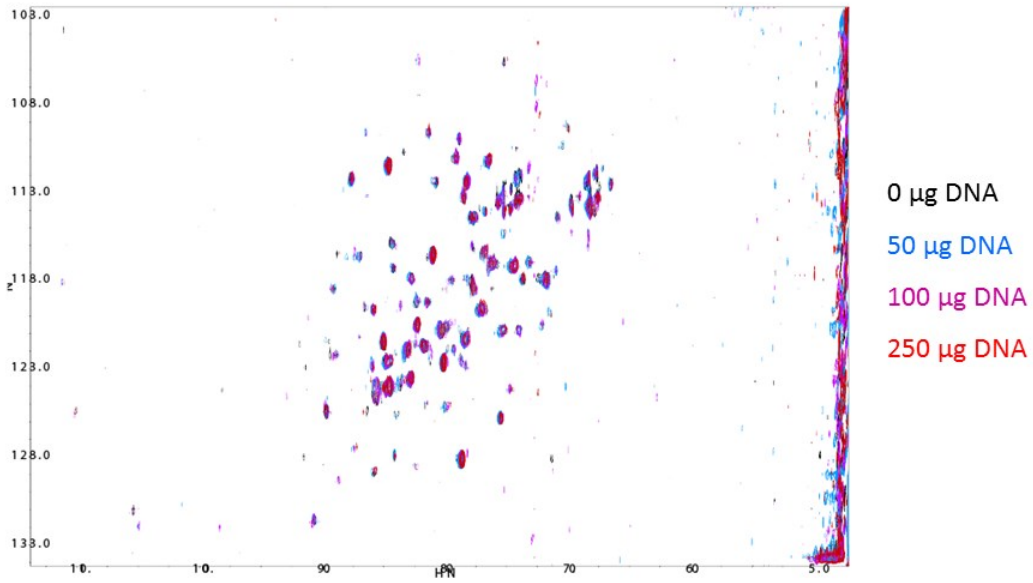


Figure 55: ^{1}H-^{15}N HSQC titration of sonicated salmon sperm DNA into CAP

Overlay of ^{1}H-^{15}N HSQC spectra collected during a titration of sonicated salmon sperm DNA into ^{15}N-labeled CAP.

Finally, CAP mixed with 250 µg sonicated salmon sperm DNA sample was mixed with a 1:1 ratio of CAP:cAMP to determine if adding cAMP to the mixture would promote binding of CAP to sonicated salmon sperm DNA (Figure 56). Adding cAMP to the mixture significantly reduced the detected signal for CAP indicating that when cAMP is present CAP will bind to nonspecific DNA thus leading to an increase in the apparent molecular weight of CAP and a decrease in the signal from CAP by NMR. This would indicate that it may be possible to detect CAP within live *E. coli* cells by NMR spectroscopy in its apo form and once CAP is bound to cAMP, and thus

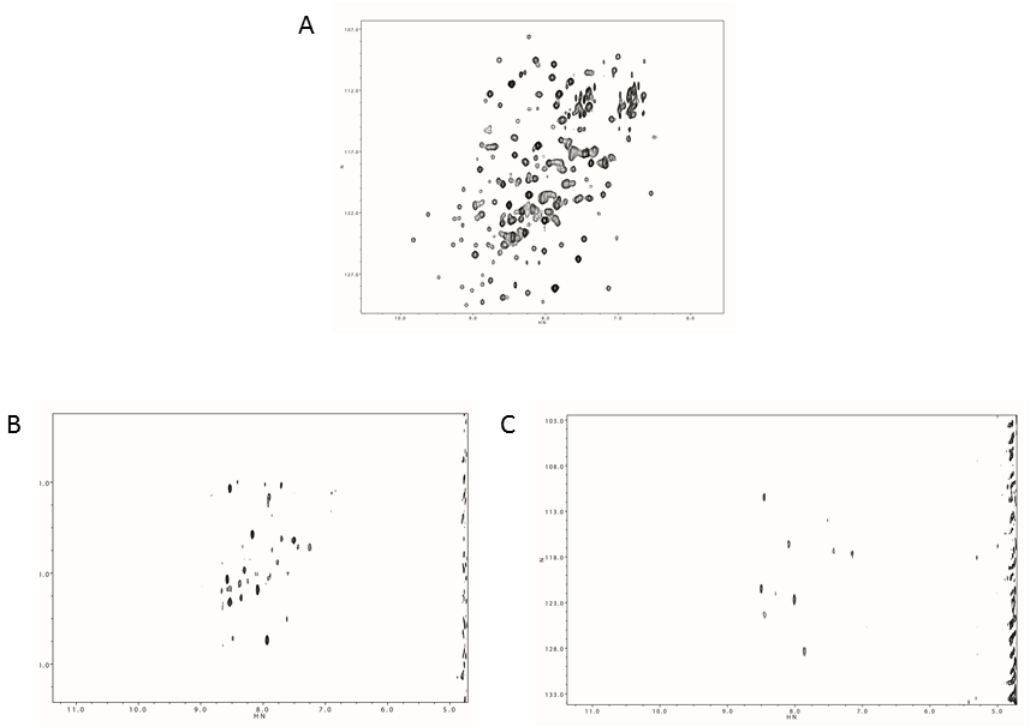


Figure 56: ^{1}H-^{15}N HSQC's of CAP bound to sonicated salmon sperm DNA with or without cAMP

^{1}H-^{15}N HSQC spectra of CAP. A) pure apo CAP B) CAP bound to 250 µg sonicated salmon sperm DNA C) 1:1 CAP: cAMP bound to 250 µg sonicated salmon sperm DNA

interacting with DNA, will no longer be detected. The final control experiment performed was to take unlabeled *E. coli* cells, lyse the cells, spin out the insoluble debris and add ^{15}N-labeled CAP to the cleared lysate. This sample was then used to collect a ^{1}H-^{15}N HSQC (Figure 57). The advantage of this experiment is that only CAP is isotopically labeled in the sample thus detection of CAP does not need to compete with

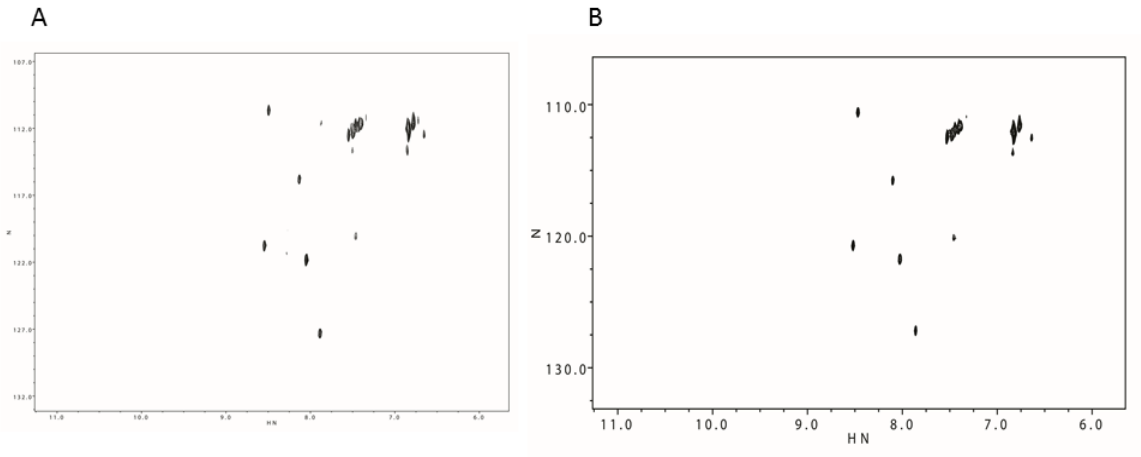


Figure 57: ^{1}H-^{15}N HSQC of CAP in unlabeled *E. coli* lysate

A) ^{1}H-^{15}N HSQC of ^{15}N-labeled CAP in clarified lysate from unlabeled *E. coli* cells. B) ^{1}H-^{15}N HSQC of ^{15}N-labeled CAP in clarified lysate from unlabeled *E. coli* cells containing DNase.

metabolites. CAP was not detected in this sample. The lack of detection may be due to non-specific DNA binding therefore deoxyribonuclase (DNase) was added to the sample to thoroughly degrade genomic DNA. This strategy was used to demonstrate that MetJ interacted with DNA in a nonspecific manner constantly *(77)*. A ^{1}H-^{15}N HSQC of the DNase degraded CAP sample in clarified lysate sample did not detect ^{15}N-labeled CAP. This would suggest that CAP is not interacting with DNA, however this particular control experiment needs further optimization before completely ruling out DNA as the reason for loss of signal. This control experiment indicated that it would not be possible to detect CAP within a living *E. coli* cell thus an in-cell NMR experiment was not performed. Further experiments, such as optimization of DNase degradation or

competition with CAPs' specific DNA recognition sequence, need to be pursued to determine why CAP is not detected in the clarified *E. coli* lysate.

Conclusions and Future Directions

In-cell NMR Spectroscopy is a technique that has been shown to be capable of detecting proteins within several different types of cells including *E. coli*, *Pichia pastoris* (*P. pastoris*), *Xenopus laevis* (*X. laevis*) oocytes and HeLa cells. While several different cell types have been used the observed proteins within these cells overlap and only about a dozen proteins have been reported to be detected to any degree by in-cell NMR spectroscopy. Many of the reported in-cell NMR protein spectra do not detect the expected number of peaks as compared to an *in vitro* spectrum but rather a subset of peaks, sometimes due to the need to use selective labeling schemes and other times due to other detection issues, such as sensitivity limitations, time constraints, apparent molecular weight changes, etc. In-cell NMR Spectroscopy has been shown to be particularly astute at using ^{1}H-^{15}N or ^{1}H-^{13}C HSQC experiments to detect changes in a protein as it binds to other proteins, drugs or DNA, however this function is only successful if a ^{1}H-^{15}N or ^{1}H-^{13}C HSQC experiment can be detected for the protein of interest. The parameters for successful protein detection within a living cell still need further understanding before this technique will be broadly applicable. The recent hypothesis that a protein's overall charge will determine the ability to detect a protein

by in-cell NMR spectroscopy is of great interest. In our observations we have shown that it is possible to detect GB1, which has a pI of 4.37, while it is not possible to detect GB1 with a six histidine c-terminal tag, which has a pI of 5.54. The pI for ubiquitin and the 3A ubiquitin we created is 6.56, quite close to the pH of the *E. coli* cytoplasm of 7.0. The pI for EGFP and mYPet is 5.80 and for CAP is 8.38. Thus for all of the proteins that we were unable to detect by in-cell NMR spectroscopy the pI is reasonably close to the pH of the *E. coli* cytoplasm. None of these proteins will have a significant overall negative charge except for GB1, which we can detect. This lack of overall negative charge may be leading to the proteins interacting via hydrophobic interactions, or in CAPs' case possibly through electrostatic interactions as well since it will be overall positively charged, with other proteins in the cytoplasm in a nonspecific manner, which will lead to an increase in the apparent molecular weight and attenuate NMR signal. To further investigate how the charge of a protein affects in-cell NMR detection the simplest experiment includes adding differently charged tags to the N and C terminus of GB1, such as a six alanine tag, a six aspartate tag and a six lysine tag. These GB1 constructs could then be examined for detection by in-cell NMR spectroscopy. It would be expected that the neutral and negative tags would not affect detection of GB1 while the positively charged tag would cause signal attenuation. Another interesting experiment would take any of these proteins, but probably start with ubiquitin, and alter the surface charge

such that the protein will present itself to the cell as a negatively charged protein regardless of its overall charge. This would indicate if overall charge truly needs to be negative or if the protein just needs to appear to be negative. In conclusion in-cell NMR Spectroscopy has the potential to be a powerful technique for studying protein structure but further gains in detecting proteins within a live cell need to be made before this technique is broadly applicable.

References

1. Franze de Fernandez, M. T., Eoyang, L., and August, J. T. (1968) Factor fraction required for the synthesis of bacteriophage Qbeta-RNA, *Nature 219*, 588-590.

2. Franze de Fernandez, M. T., Hayward, W. S., and August, J. T. (1972) Bacterial proteins required for replication of phage Q ribonucleic acid. Pruification and properties of host factor I, a ribonucleic acid-binding protein, *J Biol Chem 247*, 824-831.

3. Miranda, G., Schuppli, D., Barrera, I., Hausherr, C., Sogo, J. M., and Weber, H. (1997) Recognition of bacteriophage Qbeta plus strand RNA as a template by Qbeta replicase: role of RNA interactions mediated by ribosomal proteins S1 and host factor, *J Mol Biol 267*, 1089-1103.

4. Tsui, H. C., Leung, H. C., and Winkler, M. E. (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12, *Mol Microbiol 13*, 35-49.

5. Christiansen, J. K., Larsen, M. H., Ingmer, H., Sogaard-Andersen, L., and Kallipolitis, B. H. (2004) The RNA-binding protein Hfq of Listeria monocytogenes: role in stress tolerance and virulence, *J Bacteriol 186*, 3355-3362.

6. Vazquez-Boland, J. A., Kuhn, M., Berche, P., Chakraborty, T., Dominguez-Bernal, G., Goebel, W., Gonzalez-Zorn, B., Wehland, J., and Kreft, J. (2001) Listeria pathogenesis and molecular virulence determinants, *Clinical microbiology reviews 14*, 584-640.

7. Sobrero, P., and Valverde, C. (2011) Evidences of autoregulation of hfq expression in Sinorhizobium meliloti strain 2011, *Archives of microbiology 193*, 629-639.

8. Lybecker, M. C., Abel, C. A., Feig, A. L., and Samuels, D. S. (2010) Identification and function of the RNA chaperone Hfq in the Lyme disease spirochete Borrelia burgdorferi, *Mol Microbiol 78*, 622-635.

9. Dietrich, M., Munke, R., Gottschald, M., Ziska, E., Boettcher, J. P., Mollenkopf, H., and Friedrich, A. (2009) The effect of hfq on global gene expression and virulence in Neisseria gonorrhoeae, *FEBS J 276*, 5507-5520.

10. Brown, L., and Elliott, T. (1996) Efficient translation of the RpoS sigma factor in Salmonella typhimurium requires host factor I, an RNA-binding protein encoded by the hfq gene, *J Bacteriol 178*, 3763-3770.

11. Sonnleitner, E., Hagens, S., Rosenau, F., Wilhelm, S., Habel, A., Jager, K. E., and Blasi, U. (2003) Reduced virulence of a hfq mutant of Pseudomonas aeruginosa O1, *Microb Pathog 35*, 217-228.

12. Kulesus, R. R., Diaz-Perez, K., Slechta, E. S., Eto, D. S., and Mulvey, M. A. (2008) Impact of the RNA chaperone Hfq on the fitness and virulence potential of uropathogenic Escherichia coli, *Infect Immun 76*, 3019-3026.

13. Ding, Y., Davis, B. M., and Waldor, M. K. (2004) Hfq is essential for Vibrio cholerae virulence and downregulates sigma expression, *Mol Microbiol 53*, 345-354.

14. Sittka, A., Pfeiffer, V., Tedin, K., and Vogel, J. (2007) The RNA chaperone Hfq is essential for the virulence of Salmonella typhimurium, *Mol Microbiol 63*, 193-217.

15. Yamada, J., Yamasaki, S., Hirakawa, H., Hayashi-Nishino, M., Yamaguchi, A., and Nishino, K. (2010) Impact of the RNA chaperone Hfq on multidrug resistance in Escherichia coli, *J Antimicrob Chemother 65*, 853-858.

16. Hayashi-Nishino, M., Fukushima, A., and Nishino, K. (2012) Impact of hfq on the intrinsic drug resistance of salmonella enterica serovar typhimurium, *Front Microbiol 3*, 205.

17. Vogel, J., and Luisi, B. F. (2011) Hfq and its constellation of RNA, *Nat Rev Microbiol 9*, 578-589.

18. Frohlich, K. S., and Vogel, J. (2009) Activation of gene expression by small RNA, *Curr Opin Microbiol 12*, 674-682.

19. Soper, T., Mandin, P., Majdalani, N., Gottesman, S., and Woodson, S. A. (2010) Positive regulation by small RNAs and the role of Hfq, *Proc Natl Acad Sci U S A 107*, 9602-9607.

20. Masse, E., Escorcia, F. E., and Gottesman, S. (2003) Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli, *Genes Dev 17*, 2374-2383.

21. Morita, T., Maki, K., and Aiba, H. (2005) RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs, *Genes Dev 19*, 2176-2186.

22. Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J. C., and Vogel, J. (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation, *Nat Struct Mol Biol 16*, 840-846.

23. Mohanty, B. K., Maples, V. F., and Kushner, S. R. (2004) The Sm-like protein Hfq regulates polyadenylation dependent mRNA decay in Escherichia coli, *Mol Microbiol 54*, 905-920.

24. Hankins, J. S., Denroche, H., and Mackie, G. A. (2010) Interactions of the RNA-binding protein Hfq with cspA mRNA, encoding the major cold shock protein, *J Bacteriol 192*, 2482-2490.

25. Guisbert, E., Rhodius, V. A., Ahuja, N., Witkin, E., and Gross, C. A. (2007) Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in Escherichia coli, *J Bacteriol 189*, 1963-1973.

26. Sonnleitner, E., Schuster, M., Sorger-Domenigg, T., Greenberg, E. P., and Blasi, U. (2006) Hfq-dependent alterations of the transcriptome profile and effects on quorum sensing in Pseudomonas aeruginosa, *Mol Microbiol 59*, 1542-1558.

27. Sittka, A., Lucchini, S., Papenfort, K., Sharma, C. M., Rolle, K., Binnewies, T. T., Hinton, J. C., and Vogel, J. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq, *PLoS Genet 4*, e1000163.

28. Papenfort, K., Pfeiffer, V., Mika, F., Lucchini, S., Hinton, J. C., and Vogel, J. (2006) SigmaE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay, *Mol Microbiol 62*, 1674-1688.

29. Olejniczak, M. (2011) Despite similar binding to the Hfq protein regulatory RNAs widely differ in their competition performance, *Biochemistry 50*, 4427-4440.

30. Wagner, E. G. (2013) Cycling of RNAs on Hfq, *RNA Biol 10*, 619-626.

31. Fender, A., Elf, J., Hampel, K., Zimmermann, B., and Wagner, E. G. (2010) RNAs actively cycle on the Sm-like protein Hfq, *Genes Dev 24*, 2621-2626.

32. Zhang, A., Wassarman, K. M., Ortega, J., Steven, A. C., and Storz, G. (2002) The Sm-like Hfq protein increases OxyS RNA interaction with target mRNAs, *Mol Cell 9*, 11-22.

33. Moller, T., Franch, T., Hojrup, P., Keene, D. R., Bachinger, H. P., Brennan, R. G., and Valentin-Hansen, P. (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction, *Mol Cell 9*, 23-30.

34. Vecerek, B., Rajkowitsch, L., Sonnleitner, E., Schroeder, R., and Blasi, U. (2008) The C-terminal domain of Escherichia coli Hfq is required for regulation, *Nucleic Acids Res 36*, 133-143.

35. Olsen, A. S., Moller-Jensen, J., Brennan, R. G., and Valentin-Hansen, P. (2010) C-terminally truncated derivatives of Escherichia coli Hfq are proficient in riboregulation, *J Mol Biol 404*, 173-182.

36. Vincent, H. A., Henderson, C. A., Ragan, T. J., Garza-Garcia, A., Cary, P. D., Gowers, D. M., Malfois, M., Driscoll, P. C., Sobott, F., and Callaghan, A. J. (2012) Characterization of Vibrio cholerae Hfq provides novel insights into the role of the Hfq C-terminal region, *J Mol Biol 420*, 56-69.

37. Sun, X., Zhulin, I., and Wartell, R. M. (2002) Predicted structure and phyletic distribution of the RNA-binding protein Hfq, *Nucleic Acids Res 30*, 3662-3671.

38. Schumacher, M. A., Pearson, R. F., Moller, T., Valentin-Hansen, P., and Brennan, R. G. (2002) Structures of the pleiotropic translational regulator Hfq and an Hfq-RNA complex: a bacterial Sm-like protein, *EMBO J 21*, 3546-3556.

39. Sauter, C., Basquin, J., and Suck, D. (2003) Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli, *Nucleic Acids Res 31*, 4091-4098.

40. Beich-Frandsen, M., Vecerek, B., Sjoblom, B., Blasi, U., and Djinovic-Carugo, K. (2011) Structural analysis of full-length Hfq from Escherichia coli, *Acta Crystallogr Sect F Struct Biol Cryst Commun 67*, 536-540.

41. Wang, W., Wang, L., Zou, Y., Zhang, J., Gong, Q., Wu, J., and Shi, Y. (2011) Cooperation of Escherichia coli Hfq hexamers in DsrA binding, *Genes Dev 25*, 2106-2117.

42. Wang, W., Wang, L., Wu, J., Gong, Q., and Shi, Y. (2013) Hfq-bridged ternary complex is important for translation activation of rpoS by DsrA, *Nucleic Acids Res 41*, 5938-5948.

43. Sauer, E., and Weichenrieder, O. (2011) Structural basis for RNA 3'-end recognition by Hfq, *Proc Natl Acad Sci U S A 108*, 13065-13070.

44. Link, T. M., Valentin-Hansen, P., and Brennan, R. G. (2009) Structure of Escherichia coli Hfq bound to polyriboadenylate RNA, *Proc Natl Acad Sci U S A 106*, 19292-19297.

45. Baba, S., Someya, T., Kawai, G., Nakamura, K., and Kumasaka, T. (2010) Expression, crystallization and preliminary crystallographic analysis of RNA-binding protein Hfq (YmaH) from Bacillus subtilis in complex with an RNA aptamer, *Acta Crystallogr Sect F Struct Biol Cryst Commun 66*, 563-566.

46. Horstmann, N., Orans, J., Valentin-Hansen, P., Shelburne, S. A., 3rd, and Brennan, R. G. (2012) Structural mechanism of Staphylococcus aureus Hfq binding to an RNA A-tract, *Nucleic Acids Res 40*, 11023-11035.

47. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions, *Acta Crystallogr D Biol Crystallogr 60*, 2256-2268.

48. The PyMOL Molecular Graphics System, Version 1.5.0.4 ed., Schrödinger, LLC.

49. de Almeida Ribeiro, E., Jr., Beich-Frandsen, M., Konarev, P. V., Shang, W., Vecerek, B., Kontaxis, G., Hammerle, H., Peterlik, H., Svergun, D. I., Blasi, U., and Djinovic-Carugo, K. (2012) Structural flexibility of RNA as molecular basis for Hfq chaperone function, *Nucleic Acids Res 40*, 8072-8084.

50. Vincent, H. A., Henderson, C. A., Stone, C. M., Cary, P. D., Gowers, D. M., Sobott, F., Taylor, J. E., and Callaghan, A. J. (2012) The low-resolution solution structure of Vibrio cholerae Hfq in complex with Qrr1 sRNA, *Nucleic Acids Res 40*, 8698-8710.

51. Mikulecky, P. J., Kaw, M. K., Brescia, C. C., Takach, J. C., Sledjeski, D. D., and Feig, A. L. (2004) Escherichia coli Hfq has distinct interaction surfaces for DsrA, rpoS and poly(A) RNAs, *Nat Struct Mol Biol 11*, 1206-1214.

52. Sauer, E., Schmidt, S., and Weichenrieder, O. (2012) Small RNA binding to the lateral surface of Hfq hexamers and structural rearrangements upon mRNA target recognition, *Proc Natl Acad Sci U S A 109*, 9396-9401.

53. Jablonski, A. (1933) Efficiency of Anti-Stokes Fluorescence in Dyes, *Nature 131*, 839-840.

54. Lakowicz, J. R. (1999) *Principles of Fluorescence Spectroscopy*, kluwer Academic/Plenum Publishers, New York.

55. Lakowicz, J. R. (1992) *Topics in Fluorescence Spectroscopy - Biological Applications*, Vol. 3, Plenum Press, New York.

56. Otwinowski, Z. (1998) Denzo and Scalepackprograms: An oscillation data processing suite for macromolecular crystallography.

57. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674.

58. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.

59. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr 66*, 213-221.

60. Lundblad, J. R., Laurance, M., and Goodman, R. H. (1996) Fluorescence polarization analysis of protein-DNA and protein-protein interactions, *Mol Endocrinol 10*, 607-612.

61. Zhang, A., Schu, D. J., Tjaden, B. C., Storz, G., and Gottesman, S. (2013) Mutations in Interaction Surfaces Differentially Impact E. coli Hfq Association with Small RNAs and Their mRNA Targets, *J Mol Biol in press*.

62. Horstmann, N., Orans, J., Valentin-Hansen, P., Shelburne, S. A., 3rd, and Brennan, R. G. (2012) Structural mechanism of Staphylococcus aureus Hfq binding to an RNA A-tract, *Nucleic Acids Res*.

63. Lakowicz, J. R. (1983) *Principles of Fluorescence Spectroscopy*, Plenum, New York.

64. Tsui, H. C., Feng, G., and Winkler, M. E. (1997) Negative regulation of mutS and mutH repair gene expression by the Hfq and RpoS global regulators of Escherichia coli K-12, *J Bacteriol 179*, 7476-7487.

65. Vecerek, B., Moll, I., and Blasi, U. (2005) Translational autocontrol of the Escherichia coli hfq RNA chaperone gene, *RNA 11*, 976-984.

66. Panja, S., Schu, D. J., and Woodson, S. A. (2013) Conserved arginines on the rim of Hfq catalyze base pair formation and exchange, *Nucleic Acids Res*.

67. Dobson, C. M. (2004) Chemical space and biology, *Nature 432*, 824-828.

68. Goto, S., Okuno, Y., Hattori, M., Nishioka, T., and Kanehisa, M. (2002) LIGAND: database of chemical compounds and reactions in biological pathways, *Nucleic Acids Res 30*, 402-404.

69. Lander, E. S. e. a. (2001) Initial sequencing and analysis of the human genome, *Nature 409*, 860-921.

70. Goodsell, D. S. (1991) Inside a living cell, *Trends Biochem Sci 16*, 203-206.

71. Ellis, R. J., and Minton, A. P. (2003) Cell biology: join the crowd, *Nature 425*, 27-28.

72. Hall, D., and Minton, A. P. (2003) Macromolecular crowding: qualitative and semiquantitative successes, quantitative challenges, *Biochim Biophys Acta 1649*, 127-139.

73. Burz, D. S., Dutta, K., Cowburn, D., and Shekhtman, A. (2006) Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR), *Nat Methods 3*, 91-93.

74. Reardon, P. N., and Spicer, L. D. (2005) Multidimensional NMR spectroscopy for protein characterization and assignment inside cells, *J Am Chem Soc 127*, 10848-10849.

75. Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., Yoshimasu, M., Hayashi, N., Mikawa, T., Walchli, M., Smith, B. O., Shirakawa, M., Guntert, P., and Ito, Y. (2009) Protein structure determination in living cells by in-cell NMR spectroscopy, *Nature 458*, 102-105.

76. Ikeya, T., Sasaki, A., Sakakibara, D., Shigemitsu, Y., Hamatsu, J., Hanashima, T., Mishima, M., Yoshimasu, M., Hayashi, N., Mikawa, T., Nietlispach, D., Walchli, M., Smith, B. O., Shirakawa, M., Guntert, P., and Ito, Y. (2010) NMR protein structure determination in living E. coli cells using nonlinear sampling, *Nat Protoc 5*, 1051-1060.

77. Augustus, A. M., Reardon, P. N., and Spicer, L. D. (2009) MetJ repressor interactions with DNA probed by in-cell NMR, *Proc Natl Acad Sci U S A 106*, 5065-5069.

78. Hubbard, J. A., MacLachlan, L. K., King, G. W., Jones, J. J., and Fosberry, A. P. (2003) Nuclear magnetic resonance spectroscopy reveals the functional state of the signalling protein CheY in vivo in Escherichia coli, *Mol Microbiol 49*, 1191-1200.

79. Xie, J., Thapa, R., Reverdatto, S., Burz, D. S., and Shekhtman, A. (2009) Screening of small molecule interactor library by using in-cell NMR spectroscopy (SMILI-NMR), *J Med Chem 52*, 3516-3522.

80. Serber, Z., Keatinge-Clay, A. T., Ledwidge, R., Kelly, A. E., Miller, S. M., and Dotsch, V. (2001) High-resolution macromolecular NMR spectroscopy inside living cells, *J Am Chem Soc 123*, 2446-2447.

81. Serber, Z., Straub, W., Corsini, L., Nomura, A. M., Shimba, N., Craik, C. S., Ortiz de Montellano, P., and Dotsch, V. (2004) Methyl groups as probes for proteins and complexes in in-cell NMR experiments, *J Am Chem Soc 126*, 7119-7125.

82. Tugarinov, V., and Kay, L. E. (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods, *J Am Chem Soc 125*, 13868-13878.

83. Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C., and Kay, L. E. (1999) A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins, *J Biomol NMR 13*, 369-374.

84. Li, C., Wang, G. F., Wang, Y., Creager-Allen, R., Lutz, E. A., Scronce, H., Slade, K. M., Ruf, R. A., Mehl, R. A., and Pielak, G. J. (2010) Protein (19)F NMR in Escherichia coli, *J Am Chem Soc 132*, 321-327.

85. Wang, Q., Zhuravleva, A., and Gierasch, L. M. (2011) Exploring weak, transient protein--protein interactions in crowded in vivo environments by in-cell nuclear magnetic resonance spectroscopy, *Biochemistry 50*, 9225-9236.

86. Burz, D. S., and Shekhtman, A. (2008) In-cell biochemistry using NMR spectroscopy, *PLoS One 3*, e2571.

87. Dedmon, M. M., Patel, C. N., Young, G. B., and Pielak, G. J. (2002) FlgM gains structure in living cells, *Proc Natl Acad Sci U S A 99*, 12681-12684.

88. McNulty, B. C., Young, G. B., and Pielak, G. J. (2006) Macromolecular crowding in the Escherichia coli periplasm maintains alpha-synuclein disorder, *J Mol Biol 355*, 893-897.

89. Mandelshtam, V. A., Taylor, H. S., and Shaka, A. J. (1998) Application of the filter diagonalization method to one- and two-dimensional NMR spectra, *J Magn Reson 133*, 304-312.

90. Kupce, E., and Freeman, R. (2003) Fast multi-dimensional NMR of proteins, *J Biomol NMR 25*, 349-354.

91. Schanda, P., Kupce, E., and Brutscher, B. (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds, *J Biomol NMR 33*, 199-211.

92. Kupce, E., and Freeman, R. (2003) Projection-reconstruction of three-dimensional NMR spectra, *J Am Chem Soc 125*, 13958-13959.

93. Coggins, B. E., Venters, R. A., and Zhou, P. (2004) Generalized reconstruction of n-D NMR spectra from multiple projections: application to the 5-D HACACONH spectrum of protein G B1 domain, *J Am Chem Soc 126*, 1000-1001.

94. Coggins, B. E., and Zhou, P. (2007) Sampling of the NMR time domain along concentric rings, *J Magn Reson 184*, 207-221.

95. Barna, J. C. J., Laue, E. D., Mayger, M. R., Skilling, J., and Worrall, S. J. P. (1987) Exponential Sampling, an alternative method for sampling in two-dimensional NMR experiments, *J Magn Reson 73*, 69-77.

96. Hiller, S., Fiorito, F., Wuthrich, K., and Wider, G. (2005) Automated projection spectroscopy (APSY), *Proc Natl Acad Sci U S A 102*, 10876-10881.

97. Eghbalnia, H. R., Bahrami, A., Tonelli, M., Hallenga, K., and Markley, J. L. (2005) High-resolution iterative frequency identification for NMR as a general strategy for multidimensional data collection, *J Am Chem Soc 127*, 12528-12536.

98. Kupce, E., and Freeman, R. (2004) Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy, *J Am Chem Soc 126*, 6429-6440.

99. Venters, R. A., Coggins, B. E., Kojetin, D., Cavanagh, J., and Zhou, P. (2005) (4,2)D Projection--reconstruction experiments for protein backbone assignment: application to human carbonic anhydrase II and calbindin D(28K), *J Am Chem Soc 127*, 8785-8795.

100. Coggins, B. E., Venters, R. A., and Zhou, P. (2005) Filtered backprojection for the reconstruction of a high-resolution (4,2)D CH3-NH NOESY spectrum on a 29 kDa protein, *J Am Chem Soc 127*, 11562-11563.

101. Kupce, E., and Freeman, R. (2003) Reconstruction of the three-dimensional NMR spectrum of a protein from a set of plane projections, *J Biomol NMR 27*, 383-387.

102. Coggins, B. E., and Zhou, P. (2006) Polar Fourier transforms of radially sampled NMR data, *J Magn Reson 182*, 84-95.

103. Coggins, B. E., and Zhou, P. (2008) High resolution 4-D spectroscopy with sparse concentric shell sampling and FFT-CLEAN, *J Biomol NMR 42*, 225-239.

104. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J Biomol NMR 6*, 277-293.

105. Johnson, B. A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules, *Methods Mol Biol 278*, 313-352.

106. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T., and Clore, G. M. (1991) A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G, *Science 253*, 657-661.

107. Selenko, P., Serber, Z., Gadea, B., Ruderman, J., and Wagner, G. (2006) Quantitative NMR analysis of the protein G B1 domain in Xenopus laevis egg extracts and intact oocytes, *Proc Natl Acad Sci U S A 103*, 11904-11909.

108. Inomata, K., Ohno, A., Tochio, H., Isogai, S., Tenno, T., Nakase, I., Takeuchi, T., Futaki, S., Ito, Y., Hiroaki, H., and Shirakawa, M. (2009) High-resolution multi-dimensional NMR spectroscopy of proteins in human cells, *Nature 458*, 106-109.

109. Vijay-Kumar, S., Bugg, C. E., and Cook, W. J. (1987) Structure of ubiquitin refined at 1.8 A resolution, *J Mol Biol 194*, 531-544.

110. Bertrand, K., Reverdatto, S., Burz, D. S., Zitomer, R., and Shekhtman, A. (2012) Structure of proteins in eukaryotic compartments, *J Am Chem Soc 134*, 12798-12806.

111. Arnesano, F., Scintilla, S., Calo, V., Bonfrate, E., Ingrosso, C., Losacco, M., Pellegrino, T., Rizzarelli, E., and Natile, G. (2009) Copper-triggered aggregation of ubiquitin, *PLoS One 4*, e7052.

112. Shimomura, O., Johnson, F. H., and Saiga, Y. (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea, *Journal of cellular and comparative physiology 59*, 223-239.

113. Yang, F., Moss, L. G., and Phillips, G. N., Jr. (1996) The molecular structure of green fluorescent protein, *Nat Biotechnol 14*, 1246-1251.

114. Chalfie, M., and Kain, S. R. (2006) *Green Fluorescent Protein*, John Wiley & Sons, Inc.

115. Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. (1996) Improved green fluorescent protein by molecular evolution using DNA shuffling, *Nat Biotechnol 14*, 315-319.

116. Khan, F., Stott, K., and Jackson, S. (2003) 1H, 15N and 13C backbone assignment of the green fluorescent protein (GFP), *J Biomol NMR 26*, 281-282.

117. Popovych, N., Tzeng, S. R., Tonelli, M., Ebright, R. H., and Kalodimos, C. G. (2009) Structural basis for cAMP-mediated allosteric control of the catabolite activator protein, *Proc Natl Acad Sci U S A 106*, 6927-6932.

118. Schultz, S. C., Shields, G. C., and Steitz, T. A. (1991) Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees, *Science 253*, 1001-1007.

Biography

Kirsten Hoff was born in San Antonio, Texas on November 8th, 1985. She was born to a military family and has lived in several states and the United Kingdom. When she turned ten her parents chose to move to Maryland where her grandparents lived. Shortly after moving to Maryland her parents divorced. Kirsten and her brother were then raised by their father. Kirsten attended Colonel Zodak Magruder High School in Rockville, MD where she was a member of the 2003 graduating class. Her interest in science started in high school when she took a molecular biology class. Kirsten attended the University of Maryland: Baltimore County (UMBC) on a merit scholarship where she graduated with a Biochemistry and Molecular Biology Bachelors of Science degree in 2007. The very first day at UMBC Kirsten met her now husband Daniel Hoff. After college Kirsten entered the doctoral program at Duke University. Daniel finished his degree at UMBC and then followed Kirsten to Durham, NC where he supported her throughout her studies. They were married on May 26th, 2013. During her time at Duke University Kirsten worked in the laboratories of Dr. Leonard D. Spicer and Dr. Richard G. Brennan. While at Duke Kirsten participated in several different projects. Kirsten also received the Student Travel Stipend to attend the Experimental Nuclear Magnetic Resonance Conference in April, 2010. A current list of her publications follows:

1. Robinson, K. E., Reardon, P. N., and Spicer, L. D. (2012) In-cell NMR spectroscopy in Escherichia coli, *Methods Mol Biol 831*, 261-277.